OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: March 24, 2003, 10:05:58 ; Search time 6737 Seconds
(without alignments)
14847.312 Million cell updates/sec

Title: US-10-033-246-21
Perfect Score: 3437
Sequence: 1 caggaccaggtcttcctacg.........gaaacgacctgaactgtcaa 3437

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1455140287 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

```
 1: gb_ba:*
 2: gb_htg:*
 3: gb_in:*
 4: gb_om:*
 5: gb_ov:*
 6: gb_pat:*
 7: gb_ph:*
 8: gb_pl:*
 9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_pr:*
27: em_ro:*
28: em_sts:*
29: em_vi:*
30: em_sy:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Match Length	DB	ID	Description
1	3437	100.0	3437	6	AX468679	AX468679 Sequence
2	3431.8	99.8	4434	9	AB037861	AB037861 Homo sapi
3	3114.2	90.6	4179	9	BC004286	BC004286 Homo sapi
4	2720.4	79.2	2769	9	BC013367	BC013367 Homo sapi
5	2397.4	69.8	2418	9	BC018777	BC018777 Homo sapi
6	2203.8	64.1	2241	9	HSM800197	HSM800197 Homo sapi
7	1726	50.2	2730	10	BC010333	BC010333 Homo sapi
8	1079.4	31.4	1141	9	AL137358	AL050110 Homo sapi
9	418	12.2	149996	9	AL102953	BC010333 Mus muscu
10	194.4	5.7	144097	2	AC120290	AL137358 Homo sapi
11	186.2	5.4	219964	2	AC130221	AC102953 Homo sapi
12	107.2	3.1	144432	3	AC004642	AC120290 Rattus no
13	107.2	3.1	154840	3	AC099018	AC130221 Mus muscu
14	107.2	3.1	157851	2	AC020509	AC004642 Drosophil
15	107.2	3.1	303823	9	AE003462	AC099018 Drosophil
16	99.2	2.9	6413	3	AC116700	AC020509 Drosophil
17	94	2.7	144097	3	AC120290	AE003462 Drosophil
18	89.2	2.6	195673	2	AL807807	AC116700 Mus muscu
19	86.8	2.5	168210	2	AC116700	AC120290 Rattus no
20	85	2.5	125020	9	AF429315	AL807807 Mus muscu
21	83.6	2.4	125020	9	AF429315	AC116700 Mus muscu
22	78.4	2.3	272545	2	AC090533	AF429315 Homo sapi
23	76.4	2.2	28527	1	SCE66	AF429315 Homo sapi
24	75.2	2.2	155662	2	AC112388	AC090533 Homo sapi
25	74.2	2.2	5858	3	AF350276	SCE66 Streptomy
26	74	2.2	180418	2	AC106674	AC112388 Rattus no
27	74	2.2	227588	2	AC094571	AF350276 Nephila m
28	73.4	2.1	1150	14	HS4U1IR3	AC106674 Rattus no
29	73.4	2.1	1926	6	AX107940	AC094571 Rattus no
30	73.4	2.1	2580	6	AR109994	HS4U1IR3 Rattus no
31	73.4	2.1	5452	6	AR083151	AX107940 Homo sapi
32	73.4	2.1	5452	12	U02454	J02079 epstein-bar
33	73.4	2.1	9600	6	A92665	AX107940 Sequence
34	73.4	2.1	9600	6	AR158345	AR108994 Sequence
35	73.4	2.1	10596	6	AR158345	AR083151 Sequence
36	73.4	2.1	10737	12	I25041	U02454 Cloning vec
37	73.4	2.1	10850	12	I30503	A92665 Sequence 1
38	73.4	2.1	10850	12	XXU02428	AR158345 Sequence
39	73.4	2.1	10737	12	U02455	I25041 Sequence 15
40	73.4	2.1	184113	14	HS49558RAJ	I30503 Sequence
41	72	2.1	167636	9	AC107097	U02455 Epstein-Bar
42	71.8	2.1	49251	9	AL358216	HS49558RAJ EBV
43	71.8	2.1	172281	14	M80517	AC107097 Human DNA
44	71.4	2.1	167830	2	AC020678	AL358216 Rattus no
45	70.4	2.0	2508	9	ORAINVOL	AC020678 Homo sapi
						AL589767 Mouse DNA
						M25312 Orang-utan

ALIGNMENTS

```
RESULT 1
AX468679
LOCUS       AX468679                3437 bp    DNA     linear   PAT 16-JUL-2002
DEFINITION  Sequence 3 from Patent WO0216581.
ACCESSION   AX468679
VERSION     AX468679.1  GI:21901457
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Gao,W.Q., Polakis,P., Shou,J., Smith,V., Soriano,R., Williams,P.M.,
            Wu,T.D. and Zhang,Z.
TITLE       Compositions and methods for the diagnosis and treatment of tumor
```

```
JOURNAL      Patent: WO 0216581-A 3 28-FEB-2002;
             Genentech, Inc. (US)
FEATURES              Location/Qualifiers
source       1..3437
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
ORIGIN
BASE COUNT   558 a    1186 c    1098 g    595 t

Query Match      100.0%;   Score 3437;   DB 6;   Length 3437;
Best Local Similarity    100.0%;   Pred. No. 0;   Mismatches    0;
Matches 3437;  Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;
```

```
Qy      1 CAGGACCAAGGTCTTCCTACCCTGGAGCACCGAGGCGGGGGAGAGCCACCATGACACATCCTCGTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CAGGACCAAGGTCTTCCTACCCTGGAGCACCGAGGCGGGGGAGAGCCACCATGACACATCCTCGTG 60

Qy     61 GTCCATGCCATGGTGAATCCTGCTGACCCTGGGGCCCCGGGGGGAGACGACGACGAGTTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTCCATGCCATGGTGATCCTGCTGACCCTGGGGCCCCGGGGGGAGACGACGACGAGTTC 120

Qy    121 CAGGCGCGCTGGCTGGACAATCTGGTTTTCCGGGAGGAGAAAGCCACTGCCCACCCCTTCCCTGTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CAGGCGCGCTGGCTGGACAATCTGGTTTTCCGGGAGGAGAAAGCCACTGCCCACCCCTTCCCTGTG 180

Qy    181 GACACATCGGAGGACGCGCTGCTGCTTCCTGACTGGCTGCTGAAGCTGCGCAATGACCGTTCT 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GACACATCGGAGGACGCGCTGCTGCTTCCTGACTGGCTGCTGAAGCTGCGCAATGACCGTTCT 240

Qy    241 GAGGTGCTCCCCCTGGTTGGACGCCCCTCCCAGGACCTGCGAGCCCAGAGCTGCTGCTG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGGTGCTCCCCCTGGTTGGACGCCCCTCCCAGGACCTGCGAGCCCAGAGCTGCTGCTG 300

Qy    301 TTCGTGCAGTGCGTTGGGCATCCCCGTGTCCCAGGACCAGCATGAGCAAAACTCCTCCAGTTCCTGGAC 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTCGTGCAGTGCGTTGGGCATCCCCGTGTCCCAGGACCAGCATGAGCAAAACTCCTCCAGTTCCTGGAC 360

Qy    361 CAGGCAGTGGCCCCACGAGACTCTGGACGAGAACATCATGGACAAGAATTACATG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CAGGCAGTGGCCCCACGAGACTCTGGACGAGAACATCATGGACAAGAATTACATG 420

Qy    421 GCCCACCTGGTTGTGGAGGTCCAGTGAGAGGCCGCGGGGGGCCGGCCAGACTTTCCACTCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GCCCACCTGGTTGTGGAGGTCCAGTGAGAGGCCGCGGGGGGCCGGCCAGACTTTCCACTCC 480

Qy    481 TTGCTCACAGCCTCCCTCGCCCCCGAGACAGACGCACCCAAAACCAAAGAGC 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTGCTCACAGCCTCCCTGCCCGCCCGAGACAGACGCACCCAAAACCAAAGAGC 540

Qy    541 AGCCCAGAGCAGCCCATAGGCCAGGCCGGATTCGGGTGTGGGGACCCAGCTCCGGGTGCTG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AGCCCAGAGCAGCCCATAGGCCAGGCCGGATTCGGGTGTGGGGACCCAGCTCCGGGTGCTG 600

Qy    601 GGCCCTGAGGACGACCTGGCGGGCATGGTTCCTCCAGATTTTCCCGGTTCAGCGGGACACCCT 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GGCCCTGAGGACGACCTGGCGGGCATGGTTCCTCCAGATTTTCCCGGTTCAGCGGGACACCCT 660

Qy    661 CGGTGGCACGGAGAGCTCCAGTCCCGCCCCGGCCCTGCCCTTGCGCGGCGCAGGCCGGCCAG 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CGGTGGCACGGAGAGCTCCAGTCCCGCCCCGGCCCTGCCCTTGCGCGGCGCAGGCCGGCCAG 720

Qy    721 GAGCTGGGCCCCGGGCGTGCTCAGGGAGTGCGGTGCGTGCTGCCTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GAGCTGGGCCCCGGCGTGCCAGGGCCAGCCCCGAGGTTCCGGTGTCCTG 780

Qy    781 CAGGCCCCTCGCCACCCTGCTCAGCTCCCCACACGCGGGGTCCCTGGTCGATTCCATGCAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CAGGCCCCTCGCCACCCTGCTCAGCTCCCCACACGCGGGGTCCCTGGTCGATTCCATGCAC 840

Qy    841 CGTAGCCACTTCCTGGCCTCCCGCTCCGGCCCAGCTCTGCCAGTACCACCGCTGTGTGTG 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CGTAGCCACTTCCTGGCCTCCCGCTCCGGCCCAGCTCTGCCAGTACCACCGCTGTGTGTG 900

Qy    901 CCACAGGAGACACCGGCTTCTCTCGTCTTCCCTGGAGGGTCTCCTGACGAGGGTGG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CCACAGGAGACACCGGCTTCTCTCGTCTTCCCTGGAGGGTCTCCTGACGAGGGTGG 960

Qy    961 CTGGACAGCCCCTGGCCTTGCGGAGGGGCCCGGGCCGCACAGCTCAGGAGATCTTGCCAGC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CTGGACAGCCCCTGGCCTTGCGGAGGGGCCCGGGCCGCACAGCTCAGGAGATCTTGCCAGC 1020

Qy   1021 CAGGCCTCAGCGCCGGGGCGCCAGGCTCAGTGATGTGCCAGGCGGGTGCTCTGCCGCCGCCCAGAGTGG 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CAGGCCTCAGCGCCGGGGCGCCAGGCTCAGTGATGTGCCAGGCGGGTGCTCTGCCGCCGCCCAGAGTGG 1080

Qy   1081 GCCCTGGCCTTCCGTCAGGACCTGGAGGTGGTTCAGCTCCACCTCCGGTCCGTGCCGGGCGCC 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GCCCTGGCCTTCCGTCAGGACCTGGAGGTGGTTCAGCTCCACCTCCGGTCCGTGCCGGGCGCC 1140

Qy   1141 ACCCTGAGGTCTGGGAGCAGTCCAGTCTTGGAGCCCGGAGCCTCCAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ACCCTGAGGTCTGGGAGCAGTGCAGTCGAGCCGGAGGCCGGAGCCTCCAG 1200

Qy   1201 GGGCTGATCGAGGTGAGGTCCCCACCTGGAGAGGAGCGTGCTGACTGACTGCATTCTTCTGCC 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GGGCTGATCGAGGTGAGGTCCCCACCTGGAGAGGAGCGTGCTGACTGACTGCATTCTTCTGCC 1260

Qy   1261 ACTGCGGGATGCTGCCTCCCCCGTTTCCAGGCTGTAAGCCCGTTGTGTGGGTGAGCTCCCTG 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ACTGCGGGATGCTGCCTCCCCCGTTTCCAGGCTGTAAGCCCGTTGTGTGGGTGAGCTCCCTG 1320

Qy   1321 CTGCTGCAGGAGGAGGAGCCCGTGGGTGGGGGAAGCCGGGTGCCGGGAGCCAGCCCTG 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTGCTGCAGGAGGAGGAGCCCGTGGGTGGGGGAAGCCGGGTGCCGGGAGCCAGCCCTG 1380

Qy   1381 GAGGCCGGTGCGGCGCTGGGGGCCCTCGTCAGGGCCCTCCTAGTGGGACTGGCTGGGAAATGTGGAC 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GAGGCCGGTGCGGCGCTGGGGGCCCTCGTCAGGGCCCTCCTAGTGGGACTGGCTGGGAAATGTGGAC 1440

Qy   1441 CCCGGAGGTGGTCAGCAGCTGCCCCGCCCGACTGCAGCTCAGCGTCGCTCTTCTCCCGGAGGAAG 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CCCGGAGGTGGTCAGCAGCTGCCCCGCCCGACTGCAGCTCAGCGTCGCTCTTCTCCCGGAGGAAG 1500

Qy   1501 GGCAAAGGTCAGGCGGCCCCAGGTGCCCGTTCCGTCCGTCCCTACCTCCCCGACCCTCTTCAGGCAT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GGCAAAGGTCAGGCGGCCCAGTGCCCGTTCCGTCCGTCCCTACCTCCCCGACCCTCTTCAGGCAT 1560

Qy   1561 CAGTCCAGCGTGGCCACCACTGCACCAGTGCATCCGAGTCTTGCTCGGGCAGAGCCGGGAA 1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CAGTCCAGCGTGGCCACCACTGCACCAGTGCATCCGAGTCTTGCTCGGGCAGAGCCGGGAA 1620

Qy   1621 CAGAGGTTCGACCCCTCTGCCTCTCTGGACTTCCTGCTGCATCCATGTTCCTGCGC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CAGAGGTTCGACCCCTCTGCCTCTCTGGACTTCCTGCTGCATCCATGTTCCTGCGC 1680

Qy   1681 ATCTTGGCAGGGGCGGGGACCAGCCCACCCGCAGAACGGGGGAAGGGAGACTGGGTGGGG 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 AATCTTGGCAGGGGCGGGGACCAGCCCACCCGCAGAACGGGGGAAGGGAGACTGGGTGGGG 1740

Qy   1741 GTCTCAGGGCCCGAGAGCTCATCAGCCTGGTGGAGCTGATCCTGCGGGAGGCGGAGAGCCG 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GTCTCAGGGCCCGAGAGCTCATCAGCCTGGTGGAGCTGATCCTGCGGGAGGCGGAGAGCCG 1800

Qy   1801 AGGCCAGACGGGGACACAGCGCGCCCTGCAGCCTCAGCGGCCCTGCCCTGCTGCTC 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 AGGCCAGACGGGGACACAGCGCGCCCTGCAGCCTCAGCGGCCCTGCCCTGCTGCTC 1860

Qy   1861 AGCTGCTGTCTGTGGGGACGATGAGAGTGTCGACGAGACACCTCAGGCTGC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AGCTGCTGTCTGTGGGGACGATGAGAGTGTCGACGAGACACCTCAGGCTGC 1920

Qy   1921 ATCCAGACGTGGGGGAGCACGGTCTCGGGAAGGGCGCCGAGAGACCTTGTCGCCAGTC 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ATCCAGACGTGGGGGAGCACGGTCTCGGGAAGGGCGCCGAGAGACCTTGTCGCCAGTC 1980

Qy   1981 TACCTTACACGCCGGCCGGAGCTGCCGGTGCCTGAGGTCCTACTGCACAGGCGAAGGG 2040
```

```
Db  1981 TACTTACACGGGCCGGAGGAGCTCCGGTGTCCGGGTGCCGGTCCTACTGCACAGCGAAGGG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2041 GCTGCCAGCAGCAGCCGGTGTGCAAGCTGGACGGACGACTCATCATCACGGCTCCTT 2100
Db  2041 GCTGCCAGCAGCAGCCGGTGTGCAAGCTGGACGGACGACTCATCATCACGGCTCCTT
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2101 GCGGACACCAGCCAGCGACTCCCGGGCGGGGTTGGAGAACCAGACATGGCCTGC 2160
Db  2101 GCGGACACCAGCCAGCGACTCCCGGGCGGGGTTGGAGAACCAGACATGGCCTGC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2161 CGGAAGCTGCGGTGCGGGTGCACCCGGTCCCATTATCGCGGCG 2220
Db  2161 CGGAAGCTGCGGTGCGGGTGCACCCGGTCCCATTATCGCGGCG
         ||||||||||||||||||||||||||||||||||||||||||||
Qy  2221 CTGCCTGCAGGGCGCGCCACCCACACTCAACTTCAGGAGTTGCTCAGGAGTTCCG 2280
Db  2221 CTGCCTGCAGGGCGCGCCACCCACACTCAACTTCAGGAGTTGCTCAGGAGTTCCG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2281 TGCTTCCTGCAGCTGGGCGCGCGCTCCAGGAGTTCCGGAGCGACACCTGAGC 2340
Db  2281 TGCTTCCTGCAGCTGGGCGCGCGCTCCAGGAGTTCCGGAGCGACACCTGAGC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2341 CACCAGGGGGGCGGCTGCCTCTGTGCCCTTCATCGTGAATTACAGG 2400
Db  2341 CACCAGGGGGGCGGCTGCCTCTGTGCCCTTCATCGTGAATTACAGG
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy  2401 AAGTCCTCCCGCCATCTGGCTGCCTTCATCAACAAGTTGTCCAGTTCATCCATPAAGTAC 2460
Db  2401 AAGTCCTCCCGCCATCTGGCTGCCTTCATCAACAAGTTGTCCAGTTCATCCATPAAGTAC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2461 ATTACCTACAANTGCCCCACCAGCACGCCCATGTCCTTCCTGCAGAAGCACGCCGGCTCCAC 2520
Db  2461 ATTACCTACAANTGCCCCACCAGCACGCCCATGTCCTTCCTGCAGAAGCACGCCGGCTCCAC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2521 GACCTGTCCTTCGACAACAGTGACTGGATGCTGAAANTCCCTTCCTTCCTTCCAGGCTCAGC 2580
Db  2521 GACCTGTCCTTCGACAACAGTGACTGGATGCTGAAANTCCCTTCCTTCCTTCCAGGCTCAGC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2581 CTGCCCACCAGGAGGACGAGGACGCAGGAGGGCTGGAGAAGAGGGCCAGAGAGGAGGAGC 2640
Db  2581 CTGCCCACCAGGAGGACGAGGACGCAGGAGGGCTGGAGAAGAGGGCCAGAGAGGAGGAGC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2641 TCAGCCCGGCTCCTGCCCTCGTCCAGCGTCTCCCGTCCCGTGTTCACCCGGGCCCGGGCCGAG 2700
Db  2641 TCAGCCCGGCTCCTGCCCTCGTCCAGCGTCTCCCGTCCCGTGTTCACCCGGGCCCGGGCCGAG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2701 ATGGCCCCCTACATGAAAACGGCTTTCCCGGGGCCAAAACGGTGTGCAGTGCTGGGAGGTT 2760
Db  2701 ATGGCCCCCTACATGAAAACGGCTTTCCCGGGGCCAAAACGGTGTGCAGTGCTGGGAGGTT
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2761 CTGAGTGACATAGACGAGATGTCCCGGCGGAGACCCCAGAGTCCTGGAGGATCTGCTGGAGGTT 2820
Db  2761 CTGAGTGACATAGACGAGATGTCCCGGCGGAGACCCCAGAGTCCTGGAGGATCTGCTGGAGGTT
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2821 AACCTGCACGGGCTGTGAGGAGTCGCGCGCGGAGGTGTTGCGCGCCCAACCTGCGCCTCAGCCTG 2880
Db  2821 AACCTGCACGGGCTGATGAGCTGTTGCGCGAGAGTGTTGCGCGCCCAACCTGCGCCTCAGCCTG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2881 GCCCTGCCGCTCCAATGCAGTAACAATTGCAGCGCGGCTTTCTGCCCCCACGTTCATG 2940
Db  2881 GCCCTGCCGCTCCAATGCAGTAACAATTGCAGCGCGGCTTTCTGCCCCCACGTTCATG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2941 TACTGCCTGGGCGAGGACTTTGAGGTGGTGCACAGCGGCCTTCCGGAACCTGCCTGG 3000
Db  2941 TACTGCCTGGGCGAGGACTTTGAGGTGGTGCACAGCGGCCTTCCGGAACCTGCCTGG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3001 TACGGCTTCCTGTGTGCCAAAGACACGCGGCTGCTGTCGCTCCAGAACGGCCCTCCGGGTGGGC 3060
Db  3001 TACGGCTTCCTGTGTGCCAAAGACACGCGGCTGCTGTCGCTCCAGAACGGCCCTCCGGGTGGGC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3061 ATGTACGGCCAGAATGACCCCAGCAGCGCCCAGATCCGACGCATCCGGAGGATCCTGCATATG 3120
Db  3061 ATGTACGGCCAGAATGACCCCAGCAGCGCCCAGATCCGACGCATCCGGAGGATCCTGCATATG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3121 GAGGCCCTGGATGTGAGCTGGTGGCAGCGACCCCCTTCCAAGCCCGTCCCGTCCCGGTCC 3180
Db  3121 GAGGCCCTGGATGTGAGCTGGTGGCAGCGACCCCCTTCCAAGCCCGTCCCGTCCCGGTCC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3181 CCGGGGGATCCTTCGAGGCGAAAGCCCAGGAAGCGGTGGGGGGTTGCCTGGTCTGTCCGAGGAGGT 3240
Db  3181 CCGGGGGATCCTTCGAGGCGAAAGCCCAGGAAGCGGTGGGGGGTTGCCTGGTCTGTCCGAGGAGGT
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3241 GAGGGCGCCCAGGCCAGCAGCAGCCAGCCAGGAGCAATACTCGGAGAGCCGGCCGGGG 3300
Db  3241 GAGGGCGCCCAGGCCAGCAGCAGCCAGCCAGGAGCAATACTCGGAGAGCCGGCCGGGG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3301 CTCCGGGGCGGGGGCGGCTGCCATCAGCGGGTCCCAGCCCAGAGACCAATACTCCGACGCCCTGGGGGTGG 3360
Db  3301 CTCCGGGGCGGGGGCGGCTGCCATCAGCGGGTCCCAGCCCAGAGACCAATACTCCGACGCCCTGGGGGTGG
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3361 ACAGGCCCTGCCGCGGCGAGGCCCCGGGGATCCCCCGGGCCCGGCATGGCGCGCTGGTTTTGGTGAATGAA 3420
Db  3361 ACAGGCCCTGCCGCGGCGAGGCCCCGGGGATCCCCCGGGCCCGGCATGGCGCGCTGGTTTTGGTGAATGAA
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3421 ACGACCTGAACTGTCAA 3437
Db  3421 ACGACCTGAACTGTCAA 3437
         |||||||||||||||||
```

```
BASE COUNT    731 a   1521 c   1419 g    763 t
ORIGIN

Query Match     99.88%;  Score 3431.8;  DB 9;  Length 4434;
Best Local Similarity  99.8%;  Pred. No. 0;
Matches 3433;  Conservative 0;  Mismatches  2;  Indels  0;  Gaps  0;
```

```
Qy           1 CAGGACCAAGGTCTTCCTACGGCTGGAGGACAGCGGGGAGACAGCCACCATCCACATCCGTG 60
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1000 CAGGACCAAGGTCTTCCTACGGCTGGAGGACAGCGGGGAGACAGCCACCATCCACATCCGTG 1059
```

(sequence alignment continues — Qy/Db rows paired through to positions 1741 / 2680)

Protein translation (Qy):

YFLRLGSSQVASRVLAMKGLSIVLSEGSLRDGEEKEPPMEEDVGDVDVLQGTQWLLR
DLPRLPFDSVRSTTALAIQQAHIMEDIPQIISAVLIIYLSQHTPVEEQAQHSDLAIDV
ARIVVERSTIMSHLFSKLSPSASDAVLSALLSIFSRVRVRMRQSRKGEGEEVISWSESQ
DQVFLRWSSGEFATMEHLIVVHAMVILLIIGPPRADDSEFQALLDIWFPEEKPLPTAFL
FLDQAVAHDPQIELEONIMDKNYAHIVEVHOHERGASGGGTFHSLLIASLPPRRDSTEA
PKPKSSPEQPIGOQRIRVGTOLRVLGPEDDIAGMFAGGRLLIIFPLSPPRMWQSSSPRPVALA
LQQALGQELARVVOGSPEVPGIIVRVLQALATLLSSPEHGGALVMSMHRSFHFLACPILR
QLCQYQRQVFQDTGFSSLFLRVLLQMLOWIDSPGVEGGPLRAQIRMLASQASAGRRLS
DVRGGLIRIAELAAFRQDLEVSSTVRAVIATLRSGEOCSVEFPDLISKVLQGLIEVRS
PHLEELIFAFFSATADAASPFPACKPVPVVVSSLLGQEEEPLAGGKRPGADPGGSLPAVRL
GPSSGLLVDWLEMLDPEVVSSCPDLQLRLLFSRRKGKGQAQVPSFRPVILLTLTHQSS
WPILHQCIRVLLGKSREQRFDSAGLDIWACIIHVPRITWQGRDQRPFPQRRREELVLRV
QGPELISIVELITIAAFRTSQDGTDAACSIIQARLPLILSCCGDDESVRKTVEEHLSG
CIOOWRGSVLGRRCRDLILLQLIIQRPELRVPVPEVLLHSEGAASSSVVCKIDGLIHRTI
TLLADTSDSRALENRGADGSMACRVLAVAIPLLLLRHLPMTAALLHGRTHLNFQEFRQ
QNHLSCFLHVLGILLELLQPHVRSEHGGALWDCLLSFIRILLNYEKSSBHLAAFINKF
VQFIHKYITYNAPAAISFLQKHADPLHDLSFDNSDLVMLKSLLAGLSLPSRDDRTDRG
LDEEGEEESSAGSLPLVSVSLLFPTMEMAPYMKRLSRGQTVEDLLEVLSIDIEMSR
RRPEILSFFSTNLQRLMSSAEECCRNLAPSLALRSMONSPSIAAAFLPTFMYCLGSQD
FEVVQTALRNLPEYALILCQEHAAVLLHRAFLVGWKYGQOMDPSAQISEALRILHMEAVN"

```
Qy        1021 ...
Db        1960 CTGGACAGCCCTGGCGTGGAGGCGTGGAGGGGCAGGAGCTCAGGGATGCTTGGCCAGC 2019
```

```
Db 2740 GTCCAGGGCCCGGAGCTCATCAGCCTTGCTGGAGCTGANCCTGGCCAGGGGACGCGG 2799

Qy 1801 AGCCAGGACGGGGACACAGCCGCTCAGCCTTGCCCTGCTGCTC 1860
        |||||||||||||||||||||||||||||||||||
Db 2800 AGCCAGGACGGGGACACAGCCGCTCAGCCTCATCAGCGTTGCTGGCTCTC 2859

Qy 1861 AGCTGCTGCTGTGGGGACGATAGAGACTGGTGTCAGGAGGTTGAGGGCTGC 1920
        ||||||||||||||||||||||||||||||||||||||||||
Db 2860 AGCTGCTGCTGTGGGGACGATAGAGACTGTGGAAGGTTGAGGGCTGC 2919

Qy 1921 ATCCAGCAGTGGGAGGAGACACCTTCTCCTGCAGCTC 1980
        ||||||||||||||||||||||||||||||||
Db 2920 ATCCAGCAGTGGGAGGAGACAGCGTGCTGGAGAGGACCTTCTCCTGCAGCTC 2979

Qy 1981 TACCTACAGGGGCCGGAGCTCCACAGCCGAAGGG 2040
        ||||||||||||||||||||||||||||
Db 2980 TACCTACAGCGGCCGGGAGCTCCTACTGCACAGCGAAGG 3039

Qy 2041 GCTGCCAGCAGCACGCGTTGCAAGCTGGGACTGGAACGTGCCCATCACGTTCCTT 2100
        |||||||||||||||||||||||||||||||||||||||||||
Db 3040 GCTGCCAGCAGCACGCGTCTGCAAGCTGGACTGCCATCACGTTCCTT 3099

Qy 2101 GCGGACACCAGCAGCGACTCCCGGCGTTGGAGAACCAGCATGGCCTGC 2160
        |||||||||||||||||||||| ||||||||||||||
Db 3100 GCGGACACCAGCAGCGACTCCCGGCGTTGGAGAACCAGCATGGCCTGC 3159

Qy 2161 CGGAAGCTGCGTGGCGGCCGCCCACCCGTGCTCGCCAGCAACTCCCCATGANCGCGGGC 2220
        ||||||||||||||||||||||||||||||||||||||||||
Db 3160 CGGAAGCTGCGGTGGCCGCCCACCCGTGCTCGCCAGCAACTCCCATGATGCGCGGGC 3219

Qy 2221 CTCCTGCAGGGCCGCCACCACCCTCAACTTCCAGGAGTTCCAGGAGCAGAACCACTGAGC 2280
        ||||||||||||||||||||||||||||||||||||||||||
Db 3220 CTCCTGCAGGGCCGCCACCACCCTCAACTTCCAGGAGTTCCGGGAGCAGAACCACTGAGC 3279

Qy 2281 TGCTTCCTCTGCCACGTTGCTGCTGGGCCGTCGGAGCGCTCGCAGCGCACGTGGTGCCAG 2340
        ||||||||||||||||||||||||||||||||||||||||||
Db 3280 TGCTTCCTCCACGTTGCTGGTCGGGCCTCGCAGCCTGGAGCTGGTTCCGACCAACCTGAGC 3339

Qy 2341 CACCAGGGGGCGGCTGGTGGGACTGGCCTTCATCCGCTGCTGAATTACAGG 2400
        |||||||||||||||||||||| ||||||||||||||
Db 3340 CACCAGGGGGCGGCTGGTGGGACTGGCCTTCATCCGCTGCTGAATTACAGG 3399

Qy 2401 AAGTCCTCCCGCCCATGGCCTGCCTTCATCAACAAGTTTGTGCAGTTCAATCCATAAGTAC 2460
        ||||||||||||||||||||||||||||||||||||||||||
Db 3400 AAGTCCTCCCGCCCATGGCCTGCCTTCATCAACAAGTTTGTGCAGTTCAATCCATAAGTAC 3459

Qy 2461 ATTACCTACAAATGCCCCACCACGCCATCTCCTTCCTGACAAGCACGCGACCCGCTCCAC 2520
        |||||||||||||||||||||||||||||||||||||||||
Db 3460 ATTACCTACAAATGCCCCACCACGCCATCTCCTTCCTGACAAGCACGCGACCCGCTCCAC 3519

Qy 2521 GACCTGTCCTTCGACAACAGTGACCTGGTGATGTGAAATCCCTCCTTCGAGGGCTCAGC 2580
        ||||||||||||||||||||||||||||||||||||||||||
Db 3520 GACCTGTCCTTCGACAACAGTGACCTGGTGATGTGAAATCCCTCCTTGCCAGGGCTCAGC 3579

Qy 2581 CTTGCCCACAGGGACCACAGGACACCGACCTGGACGGAAGAGGGGCGAGGAGGAGAGAGC 2640
        ||||||||||||||||||||||||||||||||||||||||||
Db 3580 CTTGCCCACAGGGACGACAGGACGACCGAGGCCTGACGAAGAGGGGCGAGGAGGAGAGC 3639

Qy 2641 TCAGCCCGGCTCCTTGCCCCTGGTCAGCGTCCCCGTGTTCACCCCTGACCGGCCCGAG 2700
        ||||||||||||||||||||||||||||||||||||||||||
Db 3640 TCAGCCCGGCTCCTTGTCACGCGTCTCCCTGTTCACCCCTGACCGGCCCGAG 3699

Qy 2701 ATGGCCCCCTACATGAGACAGATGTCCCGCGGAGACCCCGAAGGGTTTCCGGGGGCCAAA 2760
        ||||||||||||||||||||||||||||||||||||||||||
Db 3700 ATGGCCCCCTACATGAAACGGGTTTCCGGGGTGAGGGATCTCAGCTTCTGGAGGTT 3759

Qy 2761 CTGAGTGACATAGAGACAGGAGATGTCCCGCGGGCGAGACCCGAGATCCTGAGCTTCTTCTCTGGACC 2820
        ||||||||||||||||||||||||||||||||||||||||||
Db 3760 CTGAGTGAACATAGAGACAGAGTCCCGCGGCCGAAACGGTGAGTCTGCTGGAGGTT 3819

Qy 2821 AACCTGCAGCCGCTGATGAGTGAGCTCGGCCGAGGAGAGTTGTTGCCGCCAACCTGCCCTTCGCCTTCAGCCTG 2880
        ||||||||||||||||||||||||||||||||||||||||||
Db 3820 AACCTGCAGCCGCTGATGAGCTCGGCCGAGAGTGTTGCCGCAACCTGCCCTTCAGCCTG 3879
```

```
Qy 2881 GCCCTTGCCTCCAGTCCAGAACAGACCCAGCAATTCAGCCGCGTTTCCTGCCCAGTTCATG 2940
        ||||||||||||||||||||||||||||||||||||||||||
Db 3880 GCCCTTGCGCTCCATTGCAGCCTGACCCAGCAATTCAGCCGCGTTTCCTGCCCAGTTCATG 3939

Qy 2941 TACTGCCCGGGCCAGCCAGACACTTTGAGGTCGTGCAGAGGGCCCTCCGGGAACCTGCCTGAG 3000
        ||||||||||||||||||||||||||||||||||||||||||
Db 3940 TACTGCCCGGGCCAGCCAGACACTTTGAGGTGGTGCAGAGGGCCCTCCGGGAACCTGGGC 3999

Qy 3001 TAGGCTCCTCCTGTGCCAAGACACGGGGCTGTGCTCCTCCACCGGGCGGGCCTGGCCTGGGC 3060
        ||||||||||||||||||||||||||||||||||||||||||
Db 4000 TAGGCTCCTCCTGTGCCAGACACACGGGCTGTGCTCCACCGGGCGGGCCTTCCGGAACCTGGGGC 4059

Qy 3061 ATGTACGGCCGGACGGATGGGCCCACCCGGCCAGATCTCCAGGGGCGCCGAGGATCCGAGAAGG 3120
        ||||||||||||||||||||||||||||||||||||||||||
Db 4060 ATGTACGGCCAGATGGGACCCCAGCCCACCCGGCCGACATCTCCGAGGGCCTCTGCACGCAATATG 4119

Qy 3121 GAGGCCCGTGAGTGAGGTGAGCTGTGGCCACGCCGCCGGGCCCCTTCCAGGCCCTCCGGTCCC 3180
        ||||||||||||||||||||||||||||||||||||||||||
Db 4120 GAGGCCCTGAGTGAGCTGTGGCCACGCCACCCGGCCCTCAGCCCGCCCCGTCCC 4179

Qy 3181 CCGGGGATCCTCGAGGCCAGGCCAGGAACCGGTGGGGGCGTTGCTGGTCTCGTCCGAGGAGGT 3240
        ||||||||||||||||||||||||||||||||||||||||||
Db 4180 CCGGGGATCCTCGAGGCCAAAGCCCAGGAACGTGGGGCGTTGCCGGTCTGTCTGCCGAGGAGGT 4239

Qy 3241 GAGGGCGCCCAGCCCCTGAGGCCGAGCGCCCAGGAGCAGACAATACTCCGAGGCCCTGGGGGTGG 3300
        ||||||||||||||||||||||||||||||||||||||||||
Db 4240 GAGGGCGCCCGAGCCCCTGAGGCCGAGCGCCCAGGAGCCGCGAATACTCGAGGCCTGGGGGGTGG 4299

Qy 3301 CTCCGGGCCGGCCGCTGGCCATCAGGGGCGGTCCGAGGCGCAAGCCCTCCATTCACCTTCTGGGCC 3360
        ||||||||||||||||||||||||||||||||||||||||||
Db 4300 CTCCGGGCCGGCCGCTGGCCATCAGGGGTCGTCCGAGGCCAAGCCCTCATTCACCTTCTGGGCC 4359

Qy 3361 ACAGCCCTGCCGGGAGGGAGCCGGGATCCCCCGGGCAGGGCCCAGGCGGCTGGTTTTGAATGAA 3420
        ||||||||||||||||||||||||||||||||||||||||||
Db 4360 ACAGCCCTGCCGGGAGGGACATCCCCCGGGGCATGGGCCTGGGCCTGGGTTTTGAATGAA 4419

Qy 3421 ACGAACTGAACTGTC 3435
        |||||||||||||||
Db 4420 ACGAACTGAACTGTC 4434
```

Ness, Pawan Pandoh, Anna-liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra,

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 13 Row: j Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES   Location/Qualifiers
source   1..4179
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3618123"
                     /tissue_type="Skin, melanotic melanoma."
                     /clone_lib="NIH_MGC_20"
                     /lab_host="DH10B-R"
                     /note="Vector: pOT7"
CDS   <1..2024
                     /codon_start=3
                     /product="Unknown (protein for IMAGE:3618123)"
                     /protein_id="AAH04286.1"
                     /db_xref="GI:13279125"
                     /translation="SLRLGHLLCRSRNPDFLLHIIQRQASSQSMPWLADLVQSSEGSL
                     DVLPVQCLCEFLLHDAVDDAASGEEDEGESKEGKEKRKRQRQKROQLGRLQDLLLG
                     PKADEQTICEVLDYFLRRLGSSQVASRVLAMKGLSLVLSEGSLRDGEEKPKPMEEDVG
                     DTDVLGYQWLLRDLPRLPLFDSVRSTTALALQQAIHMETDPQTISAYLIYLSQHIPV
                     EEQAQHSDLALDVARLVVERSTIMSHLFSKLSPSAASDAVLSAILSIFSRYVRMRQS
                     KEGEEVVSWSESQDQVFLRNSGSPATAMHILVVHAMVILILGPPRADSEFQAILDI
                     WFPEEKFLPTAFLVDTSEEALLPDWLKIRLRMIRSEVLRLVDAALQDLEPQQLLFVQS
                     FGIPVSMSKLLQFIDQAVAHDEPQTLEQONIMDKNYMAHIVEVQHHRGASGGQTFHSIL
                     TASLPPRDSTEAPKPKSSPEQPIGQRIRVGTQLRVGPEDDLAGMFLQIPIPLSLDP
                     RWQSSSRPVALALGQOELARVVQGSPEVPGTTVRVLQALALILSSPHGGALVMS
                     MHRSHFLACPLLRQLCQYORCVPFQODTGFSSLFLKVILQMLQWLDSPGVEGGPLRAQLR
                     MLASQASAGRLSDVRGGLLRLAEALAFRQDLEVVSSTVRFQPVSPLWW"

BASE COUNT   712 a   1412 c   1337 g   718 t
ORIGIN

Query Match   90.6%;   Score 3114.2;   DB 9;   Length 4179;
Best Local Similarity   95.4%;   Pred. No. 0;
Matches 3280; Conservative   0; Mismatches   3;   Indels 154;   Gaps   1;

Qy   1 CAGGACCAGGTCTTCCTACGCTCGGAGCACGCGGGAGACAGCCACCACATCCTTCGTG 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   867 CAGGACCAGGTCTTCCTACGCTCGGAGCACGCGGGAGACAGCCACCACATCCTTCGTG 926

Qy   61 GTCCATGCCATGGTGATCCTGCTGACGCTGGGCCCCGCCTCCGAGCCGACGACGACAGTTC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   927 GTCCATGCCATGGTGATCCTGCTGACGCTGGGCCCCGCCTCCGAGCCGACGACGACAGTTC 986

Qy   121 CAGGGCCCTGCTGCTGGACATCTGGTTCCCGGAGGAGAAGCCACTGCCGACAGCAGTTC 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   987 CAGGGCCCTGCTGCTGGACATCTGGTTCCCGGAGGAGAAGCCACTGCCGACAGCAGTTC 1046

Qy   181 GACACATCCGGAGGAGGGGCCTGCTTCCTGACTGGGCTCCATCCGGTTCT 240
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1047 GACACATCCGGAGGAGGGGCCTGCTTCCTGACTGGGCTCCATCCGGTTCT 1106

Qy   241 GAGGTGCTCCGCGCTGGTGGACGCGCGCCCCCTGCAGGAGACCTGGAGCAGCTGCTGCTG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1107 GAGGTGCTCCGCGCTGGTGGACGCGCGCCCCCTGCAGGAGACCTGGAGCAGCTGCTGCTG 1166

Qy   301 TTCGTGCAGTCGTTTTGGCATCCCCGGTGTCCAGCATGAGAAACTCTCCAGTTCCTGGAC 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1167 TTCGTGCAGTCGTTTTGGCATCCCCGGTGTCCAGCATGAGAAACTCTCCAGTTCCTGGAC 1226

Qy   361 CAGGCAGTGGCCCACAGACTCTGGAGCAGAACATCATGACAAGAATTACATG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1227 CAGGCAGTGGCCCACAGACTCTGGAGCAGAACATCATGACAAGAATTACATG 1286

Qy   421 GCCCACCTGGGTGTGGAGGTCCAGCATGAGCGCGGCGGCGGCCTTCGGGAGGCCAGAGACTTTCCACTCC 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1287 GCCCACCTGGGTGTGGAGGTCCAGCATGAGCGCGGCGGCGGCCTTCGGGAGGCCAGAGACTTTCCACTCC 1346

Qy   481 TTGCTCACAGCCTCCCTCCCTGCCCGCCCCGCCCCGGCCGAGCACGACACAGAGGCCACCAAAACCAAAGAGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1347 TTGCTCACAGCCTCCCTCCCTGCCCGCCCCGCCCCGGCCGAGCACGACACAGAGGCCACCAAAACCAAAGAGC 1406

Qy   541 AGCCCAGACAGCCAGGGCCCATAGGCCCCAGGGCCCGGACATCGGGTGGGAGGGGGACCAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1407 AGCCCAGACAGCCAGGGCCCATAGGCCCCAGGGCCCGGACATCGGGTGGGAGGGGGACCAG 1466

Qy   601 GGCCCTGAGGACCACTCGGGCTGGGGCCATGAGTTCCTCCAGATTTTCCCGCTCAGCCCTCCCT 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1467 GGCCCTGAGGACCACTCGGGCTGGGGCCATGAGTTCCTCCAGATTTTCCCGCTCAGCCCTCCCT 1526

Qy   661 CGGTGGCAGAGCTCCAGTCCCGCGTGGCCCTTGGCCCTCGAGCAGGGCCCTGGGCCAG 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1527 CGGTGGCAGAGCTCCAGTCCCGCGTGGCCCTTGGCCCTCGAGCAGGGCCCTGGGCCAG 1586

Qy   721 GAGCTGGCCCGGGGCGTCTCAGGGGCAGGCCCGGCCGCCGGGCGATCACGGTGCGTGCCTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1587 GAGCTGGCCCGGGGCGTCTCAGGGGCAGGCCCGGCCGCCGGGCGATCACGGTGCGTGCCTG 1646

Qy   781 CAGGGCCCTCGCCCACCCTGCTCCAGGTGCCAGCCCCGGAGGCCCCGGCAGCCCGCGGGGCGGGGCCACAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1647 CAGGGCCCTCGCCCACCCTGCTCCAGGTGCCAGCCCCGGAGGCCCCGGCAGCCCGCGGGGCGGGGCCACAC 1706

Qy   841 CGTAGCCACTCCCTGCCCTGCCCGGTGGTCAGCACGCGCTCACCCGT 900
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1707 CGTAGCCACTCCCTGCCCTGCCCGGTGGTCAGCACGCGCTCACCCGT 1766

Qy   901 CCACAGGACACCGGGTTCTGCTTCCTCCGCTGCTCCTCCGCAGCGACGATGCTTGCAGTGG 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1767 CCACAGGACACCGGGTTCTGCTTCCTCCGCTGCTCCTCCGCAGCGACGATGCTTGCAGTGG 1826

Qy   961 CTGACACGCCTGGCCGTCGGAGGCGCCGGCCCACAGCCTCAGGAGATGCTTGCCAGC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1827 CTGACACGCCTGGCCGTCGGAGGCGCCGGCCCACAGCCTCAGGAGATGCTTGCCAGC 1886

Qy   1021 CAGGCCCTCAGCCGTGGGCAGGGCTCCAGGTGAGTGCGGAGGGGGCTCCTGCGCCTGGCCGAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1887 CAGGCCCTCAGCCGTGGGCAGGGCTCCAGGTGAGTGCGGAGGGGGCTCCTGCGCCTGGCCGAG 1946

Qy   1081 GCCCTGCCTTCCGTCAGGACCTGGAGGTGGTCAGCGTCCCACCGTCCGTCGCGGTCCATCGAGCC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1947 GCCCTGCCTTCCGTCAGGACCTGGAGGTGGTCAGCGTCCCACCGTCCGTCGCGGTCCATCGAGCC 1990

Qy   1141 ACCCTGAGGTCTGGGAGCAGTGCAGCGTGGAGCCGGACGTGGTGTG 1200
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1991 ACCCTGAGGTCTGGGAGCAGTGCAGCGTGGAGCCGGACGTGGTGTG 1990

Qy   1201 GGGCTGATCGAGGTGAGGTGAGGTCCCCCACCTGGAGGAGGTGCTGACTGCATTCTTCTGCC 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1991 GGGCTGATCGAGGTGAGGTGAGGTCCCCCACCTGGAGGAGGTGCTGACTGCATTCTTCTGCC 1990

Qy   1261 ACTGCGGATGCCTGCCCGTTCCCAGTTTGTAAGCCGCGTTGTGGGTGGGGTGAGCTCCCTG 1320
          -------------||||||||||||||||||||||||||||||||||||||||||||||||
Db   1991 ------------CCGTTCCCAGCCTCCCAGTTTGTAAGCCGCGTTGTGGGTGGGGTGAGCTCCCTG 1990

Qy   1321 CTTGCTCGCAGGAGGAAGAGGCCCCTGCCTGGGGGGGGGAAGCCCCGTTGTGGGGTCGGGGCAGCCTG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2033 CTTGCTCGCAGGAGGAAGAGGCCCCTGCCTGGGGGGGGGAAGCCCCGTTGTGGGGTCGGGGCAGCCTG 2092

Qy   1381 GAGGCCGTGGCTGCAGCACGTCAGCCGCCTCCCTAGTGGACTGGGCCTGGGCGGGTGGCAGCCTG 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2093 GAGGCCGTGGCTGCAGCACGTCAGCCGCCTCCCTAGTGGACTGGGCCTGGGCGGGTGGCAGCCTG 2152

Qy   1441 CCCGAGGTGGTCAGCAGCTGCCCCGACCTGCAGCTGCTCTGTCTCTTCCCGGAGGAAG 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2153 CCCGAGGTGGTCAGCAGCTGCCCCGACCTGCAGCTGCTCTGTCTCTTCCCGGAGGAAG 2212

Qy   1501 GGCAAAGGTCAGCCAGGTGTCCCTCGTTCCCTACCTCCTGACCCGTTCTTCACGGCAT 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db   2213 GGCAAAGGTCAGGCCCAGGTGCCCTCGTTCCGTCCCTACCTCCTGACCCTCTTCACGCAT 2272

Qy   1561 CAGTCCAGCTGGCCCACACTGCACCAGTGCATCCGAGTCCTGCTGGGCAAGAGCCGGGAA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2273 CAGTCCAGCTGGCCCACACTGCACCAGTGCATCCGAGTCCTGCTGGGCAAGAGCCGGGAA 2332

Qy   1621 CAGAGGTTCGACCCCTCTGCCTCTCTGGACTTCCTCTGGGCCTGCATCCATGTTCCTCGC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2333 CAGAGGTTCGACCCCTCTGCCTCTCTGGACTTCCTCTGGGCCTGCATCCATGTTCCTCGC 2392

Qy   1681 ATCTGGCAGGGGCGGGACCAGCGCACCCCGCAGAAGCGGCGGGAGGAGCTGGTGCTGCGG 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2393 ATCTGGCAGGGGCGGGACCAGCGCACCCCGCAGAAGCGGCGGGAGGAGCTGGTGCTGCGG 2452

Qy   1741 GTCCAGGGCCCGGAGCTCATCAGCCTGGTGGAGCTGATCCTGGCCGAGGCGGAGACGCGG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2453 GTCCAGGGCCCGGAGCTCATCAGCCTGGTGGAGCTGATCCTGGCCGAGGCGGAGACGCGG 2512

Qy   1801 AGCCAGGACGGGGACACAGCCGCCTGCAGCCTCATCCAGGCCCGGCTGCCCCTGCTGCTC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2513 AGCCAGGACGGGGACACAGCCGCCTGCAGCCTCATCCAGGCCCGGCTGCCCCTGCTGCTC 2572

Qy   1861 AGCTGCTGCTGTGGGGACGATGAGAGTGTCAGGAAGGTGACGGAGCACCTGTCAGGCTGC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2573 AGCTGCTGCTGTGGGGACGATGAGAGTGTCAGGAAGGTGACGGAGCACCTGTCAGGCTGC 2632

Qy   1921 ATCCAGCAGTGGGGAGACAGCGTGCTGGGAAGGCGCTGCCGAGACCTTCTCCTGCAGCTC 1980
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   2633 ATCCAGCAGTGGGGAGACAGCGTGCTGGGCAGGCGCTGCCGAGACCTTCTCCTGCAGCTC 2692

Qy   1981 TACCTACAGCGGCCGGAGCTGCGGGTGCCCGTGCCTGAGGTCCTACTGCACAGCGAAGGG 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2693 TACCTACAGCGGCCGGAGCTGCGGGTGCCCGTGCCTGAGGTCCTACTGCACAGCGAAGGG 2752

Qy   2041 GCTGCCAGCAGCAGCGTCTGCAAGCTGGACGGACTCATCCACCGCTTCATCACGCTCCTT 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2753 GCTGCCAGCAGCAGCGTCTGCAAGCTGGACGGACTCATCCACCGCTTCATCACGCTCCTT 2812

Qy   2101 GCGGACACCAGCGACTCCCGGGCGTTGGAGAACCGAGGGGCGGATGCCAGCATGGCCTGC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2813 GCGGACACCAGCGACTCCCGGGCGTTGGAGAACCGAGGGGCGGATGCCAGCATGGCCTGC 2872

Qy   2161 CGGAAGCTGGCGGTGGCGCACCCGCTGCTGCTGCTCAGGCACCTGCCCATGATCGCGGCG 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2873 CGGAAGCTGGCGGTGGCGCACCCGCTGCTGCTGCTCAGGCACCTGCCCATGATCGCGGCG 2932

Qy   2221 CTCCTGCACGGCCGCACCCACCTCAACTTCCAGGAGTTCCGGCAGCAGAACCACCTGAGC 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2933 CTCCTGCACGGCCGCACCCACCTCAACTTCCAGGAGTTCCGGCAGCAGAACCACCTGAGC 2992

Qy   2281 TGCTTCCTGCACGTGCTGGGCCTGCTGGAGCTGCTGCAGCCGCACGTGTTCCGCAGCGAG 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2993 TGCTTCCTGCACGTGCTGGGCCTGCTGGAGCTGCTGCAGCCGCACGTGTTCCGCAGCGAG 3052

Qy   2341 CACCAGGGGGCGCTGTGGGACTGCCTTCTGTCCTTCATCCGCCTGCTGCTGAATTACAGG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3053 CACCAGGGGGCGCTGTGGGACTGCCTTCTGTCCTTCATCCGCCTGCTGCTGAATTACAGG 3112

Qy   2401 AAGTCCTCCCGCCATCTGGCTGCCTTCATCAACAAGTTTGTGCAGTTCATCCATAAGTAC 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3113 AAGTCCTCCCGCCATCTGGCTGCCTTCATCAACAAGTTTGTGCAGTTCATCCATAAGTAC 3172

Qy   2461 ATTACCTACAATGCCCCAGCAGCCATCTCCTTCCTGCAGAAGCACGCCGACCCGCTCCAC 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3173 ATTACCTACAATGCCCCAGCAGCCATCTCCTTCCTGCAGAAGCACGCCGACCCGCTCCAC 3232

Qy   2521 GACCTGTCCTTCGACAACAGTGACCTGGTGATGCTGAAATCCCTCCTTGCAGGGCTCAGC 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3233 GACCTGTCCTTCGACAACAGTGACCTGGTGATGCTGAAATCCCTCCTTGCAGGGCTCAGC 3292

Qy   2581 CTGCCCAGCAGGGACGACAGGACCGACCGAGGCCTGGACGAAGAGGGCGAGGAGGAGAGC 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3293 CTGCCCAGCAGGGACGACAGGACCGACCGAGGCCTGGACGAAGAGGGCGAGGAGGAGAGC 3352
```

```
Qy   2641 TCAGCCGGCTCCTTGCCCCTGGTCAGCGTCTCCCTGTTCACCCCTCTGACCGCGGCCGAG 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3353 TCAGCCGGCTCCTTGCCCCTGGTCAGCGTCTCCCTGTTCACCCCTCTGACCGCGGCCGAG 3412

Qy   2701 ATGGCCCCCTACATGAAACGGCTTTCCCGGGGCCAAACGGTGGAGGATCTGCTGGAGGTT 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3413 ATGGCCCCCTACATGAAACGGCTTTCCCGGGGCCAAACGGTGGAGGATCTGCTGGAGGTT 3472

Qy   2761 CTGAGTGACATAGACGAGATGTCCCGGCGGAGACCCGAGATCCTGAGCTTCTTCTCGACC 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3473 CTGAGTGACATAGACGAGATGTCCCGGCGGAGACCCGAGATCCTGAGCTTCTTCTCGACC 3532

Qy   2821 AACCTGCAGCGGCTGATGAGCTCGGCCGAGGAGTGTTGCCGCAACCTCGCCTTCAGCCTG 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3533 AACCTGCAGCGGCTGATGAGCTCGGCCGAGGAGTGTTGCCGCAACCTCGCCTTCAGCCTG 3592

Qy   2881 GCCCTGCGCTCCATGCAGAACAGCCCCAGCATTGCAGCCGCTTTCCTGCCCACGTTCATG 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3593 GCCCTGCGCTCCATGCAGAACAGCCCCAGCATTGCAGCCGCTTTCCTGCCCACGTTCATG 3652

Qy   2941 TACTGCCTGGGCAGCCAGGACTTTGAGGTGGTGCAGACGGCCCTCCGGAACCTGCCTGAG 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3653 TACTGCCTGGGCAGCCAGGACTTTGAGGTGGTGCAGACGGCCCTCCGGAACCTGCCTGAG 3712

Qy   3001 TACGCTCTCCTGTGCCAAGAGCACGCGGCTGTGCTGCTCCACCGGGCCTTCCTGGTGGGC 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3713 TACGCTCTCCTGTGCCAAGAGCACGCGGCTGTGCTGCTCCACCGGGCCTTCCTGGTGGGC 3772

Qy   3061 ATGTACGGCCAGATGGACCCCAGCGCGCAGATCTCCGAGGCCCTGAGGATCCTGCATATG 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3773 ATGTACGGCCAGATGGACCCCAGCGCGCAGATCTCCGAGGCCCTGAGGATCCTGCATATG 3832

Qy   3121 GAGGCCGTGATGTGAGCCTGTGGCAGCCGACCCCCCTCCAAGCCCCGGCCCGTCCCGTCC 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3833 GAGGCCGTGATGTGAGCCTGTGGCAGCCGACCCCCCTCCAAGCCCCGGCCCGTCCCGTCC 3892

Qy   3181 CCGGGGATCCTCGAGGCAAAGCCCAGGAAGCGTGGGCGTTGCTGGTCTGTCCGAGGAGGT 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3893 CCGGGGATCCTCGAGGCAAAGCCCAGGAAGCGTGGGCGTTGCTGGTCTGTCCGAGGAGGT 3952

Qy   3241 GAGGGCGCCGAGCCCTGAGGCCAGGCAGGCCCAGGAGCAATACTCCGAGCCCTGGGGTGG 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3953 GAGGGCGCCGAGCCCTGAGGCCAGGCAGGCCCAGGAGCAATACTCCGAGCCCTGGGGTGG 4012

Qy   3301 CTCCGGGCCGGCCGCTGGCATCAGGGGCCGTCCAGCAAGCCCTCATTCACCTTCTGGGCC 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4013 CTCCGGGCCGGCCGCTGGCATCAGGGGCCGTCCAGCAAGCCCTCATTCACCTTCTGGGCC 4072

Qy   3361 ACAGCCCTGCCGCGGAGCGGCGGATCCCCCCGGGCATGGCCTGGGCTGGTTTTGAATGAA 3420
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4073 ACAGCCCTGCCACGGAGCGGCGGATCCCCCCGGGCATGGCCTGGGCTGGTTTTGAATGAA 4132

Qy   3421 ACGACCTGAACTGTCAA 3437
          |||||||||||||||||
Db   4133 ACGACCTGAACTGTCAA 4149
```

```
RESULT 4
BC013367
LOCUS       BC013367                2769 bp    mRNA    linear   PRI 04-SEP-2001
DEFINITION  Homo sapiens, clone IMAGE:3677373, mRNA, partial cds.
ACCESSION   BC013367
VERSION     BC013367.1  GI:15426515
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2769)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-AUG-2001) National Institutes of Health, Mammalian
```

```
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Institute for Systems Biology
          http://www.systemsbiology.org
          contact: amadan@systemsbiology.org
          Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
          Greene, Mark Ketteman and Anuradha Madan
          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAL Plate: 24 Row: m Column: 10
          This clone was selected for full length sequencing because it
          passed the following selection criteria: Hexamer frequency ORF
          analysis.
FEATURES             Location/Qualifiers
     source          1..2769
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3677373"
                     /tissue_type="lymph, Burkitt lymphoma"
                     /clone_lib="NIH_MGC_8"
                     /lab_host="DH10B-R"
                     /note="vector: pOT87"
     CDS             <1..2439
                     /codon_start=1
                     /product="Unknown (protein for IMAGE:3677373)"
                     /protein_id="AAH13367.1"
                     /db_xref="GI:15426516"
                     /translation="GTRQALGQHLARVVQGSPEYPGITIVRVLQALATILLSSPHGGALV
                     MSMHRSHFLACPLIRQLCQYORCVPQDTPGHSSLFLKVLLQMIQWIDSPGVEGGPLRAQ
                     DLISKVLQGLIEVRSPHLEELILAFFSATDADAASPPFACKPVVVWVSSILLQEEEPLAG
                     GKPGADGGSLEAVRLGPSSGLLVDWLEMDDPEVVSSCPDLQLRLFSRKRGKQAOVP
                     SFRPYLTLFTHQSSWPILHQCIRVLLGKSREQKFDPSASIDFLWRACIHVPRIWQGRD
                     QRPPOKRREELVLRVQGPELILSLVELIILAPAEFRSQDGPDAACSLIQARLPILLSCC
                     GDDESVRKVTEHLSCCIQOWGDSVLGRRCHDLIILQIVLQRPELRVPVPEVLLHSEGAA
                     YRKSRHLAPINKRVQFIHKYIIWNAPAAISFLQKHADPLHDLSFDNSDLVMLKSLL
                     SSSVCKLDGLIHRIITLLADYSDSRALENRGADASMACRKLAVAHPLILIIRHLPMIAA
                     EDLLEVLSDIDEMSRRPBILSFFSTNLQRLMSSAEQGLALLMDCLLSFIITLIN
                     AAPLPTFMYCLGSQDFEVVQTALRNLPEYALLCQEHAAVLLHRAFLVGMYGQOMDPSAQ
                     ISEALRILHEAAVM"
BASE COUNT     462 a    940 c    891 g    476 t
ORIGIN

Query Match     79.28;    Score 2720.4;    DB 9;    Length 2769;
Best Local Similarity   99.9%;   Pred. No. 0;
Matches 2732;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   1;

Qy    705 GCAGCGCCTGGGCCGAGGAGCTGGCCCGCGCTCGTCAGGGCGCCCAGGCCCAGGCGCAT 764
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9 GCAGCGCCTGGGCCGAGGAGCTGGCCCGCGCTCGTCAGGGCGCCCAGGCCCAGGCGCAT 68

Qy    765 CACGGTGCCGTTGCCTCCGCCGCACCCTGCTCAGCTCCCCACACGCGCGGTTGCCCCT 824
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69 CACGGTGCCGTTGTCCTGCGAGGCCCCTGCCTCAGCTCCCCACACGCGCGGTTGCCCCT 128

Qy    825 GGTGATGTGTCCATGCACCGTAGCCACTTCCTGCCCTGGCCCTGCGCCAGCTCTGCCA 884
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    129 GGTGATGTGTCCATGCACCGTAGCCACTTCCTGCCCTGGCCCTGCGCCAGCTCTGCCA 188

Qy    885 GTACCAGCGCGCTGTGTGCCACAGGACACCGGGCTTCTCCCTCGCTCTTCCTGAAGGTGCCT 944
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    189 GTACCAGCGCCTGTGTGTGCCACAGAGACACCGGGCTTCTCCCTCGCTCTTCCTGAAGGTGCCTCCT 248

Qy    945 GCAGAGATGCTCCAGTGGGCTGGGTGGGACAGCCCCTGCGTGGAGGGGCCCTGGGCGCAGCT 1004
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    249 GCAGAGATGCTCCAGTGGGCTGGGTGGGACAGCCCCTGCGTGGAGGGGCCCTGGGCACAGCT 308

Qy   1005 CAGGAGCACTTGCCAGCCAGGCCCTCAGCCGTGCGCAGCCCTCAGTGATGTGCCAGGGCT 1064
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    309 CAGGAGCACTTGCCAGCCAGGCCCTCAGCCGTGCGCAGCCCTCAGTGATGTGCCAGGGCT 368

Qy   1065 CCTGCGCCCTGGCCGGAGCCCTGGGCTTCCCGTCCAGGAGCCTGCAGGGACTCCACCGT 1124
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369 CCTGCGCCCTGGCCGGAGCCCTGGGCTTCCCGTCCAGGAGCCTGCAGGGACTCCACCGT 428

Qy   1125 CCGTGCCGGTCATGCCACCCCTGAGGTCTGGGGAGCAGTGCAGCGTGGAGCCACCTGAT 1184
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 CCGTGCCGGTCATGCCACCCCTGAGGTCTGGGGAGCAGTGCAGCGTGGAGCCACCTGAT 488

Qy   1185 CAGCAAAGTCCTCCAGGGGTGATCGACGGGAGGTGAGGTCCCCCCCCCACCTGGAGGAGCGTGCTGAC 1244
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 CAGCAAAGTCCTCCAGGGGCTGATCGACGGGAGGTGAGGTCCCCCCACCTGGAGGAGCGTGCTGAC 548

Qy   1245 TGCATTCTTCTGCCACTCGCGATGCTGCCCCCGTTTCCAGCCTGTGTAAGCCCGTTGT 1304
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    549 TGCATTCTTCTGCCACTCGCGATGCTGCCCCCGTTTCCAGCCTGTGTAAGCCCGTTGT 608

Qy   1305 GGTGGTGAGCTCCCCGCCTGCCTGCGGAGGAGGAGCCCCTGGCCGGGGCGGAGGCGTGC 1364
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 GGTGGTGAGCTCCCCGCCTGCCTGCGGAGGAGGAGCCCCTGGCCGGGGCGGAGGCGTGC 668

Qy   1365 GGACGGTGGCAGCCTGGAGGCCGTCTGGAGCCTGGGCTGGTCAGGCCTCTTCAGTGGACTG 1424
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669 GGACGGTGGCAGCCTGGAGGCCGTCTGGAGCCTGGGCTGGTCAGGCCTCTTCAGTGGACTG 728

Qy   1425 GCTGGAAATGCTGGACCCCGGAGTGGTCACGAGCTGCCCCGAGGTGGTCACGAGCTCAGGCTGCT 1484
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729 GCTGGAAATGCTGGACCCCGGAGTGGTCACGAGCTGCCCCGAGGTGGTCACGAGCTCAGGCTGCT 788

Qy   1485 CTTCTCCCGGAGGAAGGGCAAAGGTCAGGGCGCCTCGTTCGGTCCCTTACCTCCT 1544
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    789 CTTCTCCCGGAGGAAGGGCAAAGGTCAGGGCGCCTCGTTCGGTCCCTTACCTCCT 848

Qy   1545 GACCCTCTTCACGCAATCAGTCAGTGGCCCCACAGTGCCATCCGAGTCCTGCT 1604
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    849 GACCCTCTTCACGCAATCAGTCAGTGGCCCCACAGTGCCATCCGAGTCCTGCT 908

Qy   1605 GGGCAAGAGCCGGGAACAGAGGTTCGACCCCTGCTGGACTTCCTCTTGGGCCTG 1664
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    909 GGGCAAGAGCCGGGAACAGAGGTTCGACCCCTGCTGGACTTCCTCTTGGGCCTG 968

Qy   1665 CATCCAAGTTCCTGCATCTGCAGGGCGCGGGACCACCCCGCAGAAGCGGCGGGA 1724
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    969 CATCCAAGTTCCTGCATCTGCAGGGCGCGGGACCACCCCGCAGAAGCGGCGGGA 1028

Qy   1725 GGAGCTGGTGCTGCGGGTTCAGGGGCCGGAGCTCATCAGCCTGGTGGAGCTGGC 1784
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1029 GGAGCTGGTGCTGCGGGTTCAGGGGCCGGAGCTCATCAGCCTGGTGGAGCTGGC 1088

Qy   1785 CGAGGCGGAGAGCGGGACCAGGAGCAGCGGGACAGCCTTCATCAGCCCG 1844
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 CGAGGCGGAGAGCGGGACCAGGAGCAGCGGGACAGCCTTCATCAGCCCG 1148

Qy   1845 GCTGCCCCGCTGCTCTGCTCAGCTGCTCGTCTGTGGGACGATGAGAGTGTCAGGAAGGTGACGGA 1904
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149 GCTGCCCCGCTGCTCTGCTCAGCTGCTCGTCTGTGGGACGATGAGAGTGTCAGGAAGGTGACGGA 1208

Qy   1905 GCACCTGTCAGGGTGCGATCCAGCAGTGGGGAGACAACCGTGCTGGAAGGCCTTCGCCAGA 1964
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209 GCACCTGTCAGGGTGCGATCCAGCAGTGGGGAGACAACCGTGCTGGAAGGCCTTCGCCAGA 1268

Qy   1965 CCTTCTCCTGCAGCTCTTACCTCACAACGCGCCGGAGCCTGGGGTCGCCCCGCTGAGGTCCT 2024
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 CCTTCTCCTGCAGCTCTTACCTCACAACGCGCCGGAGCCTGGGGTCGCCCCGCTGAGGTCCT 1328

Qy   2025 ACTGCACACGGAAGGGGCTCCCAGCAGCAGCGGTCTGCAAGCTGGACGGACGACTCATCACACG 2084
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1329 ... 1328
```

```
Db  1329 ACTGCACAGCGGAAGGGGCGGCTGCCAGCCAGCCAGGCGTTCTGCAAGCTGGACGGACTCCATCCACG 1388
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2085 CTTCATCACGCTCCTTGCGGACACAGCAGCGGCGATCCAGCCAGCCAGGCGTTCTGCAAGCTGGACGGACTCCATCCACG 2144

Db  1389 CTTCATCACGCTCCTTGCGGACACAGCAGCGGCGA 1448
         |||||||||||||||||||||||||||||||||||
Qy  2145 TGCCAGCCATGGCCTGCTGCGGAAGCTGGTGGGGA 2204

Db  1449 TGCCAGCCATGGCCTGCTGCGGAAGCTGGTGGGGA 1508
         |||||||||||||||||||||||||||||||||||
Qy  2205 GCCCATGATCGCGCGCGGCGCGCCACCCACTCAACTTCCAGGAGGTTCCGGCA 2264

Db  1509 GCCCATGATCGCGCGCGGCGCGCCACCCACTCAACTTCCAGGAGGTTCCGGCA 1568
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2265 GCAGAACCACCTGAGCTGCTTCCTGCCACGGGCCGCGCACCACTCAACTTCCAGGAGGTTCCGGCA 2324

Db  1569 GCAGAACCACCTGAGCTGCTTCCTGCCACGGGCCG 1628
         |||||||||||||||||||||||||||||||||||
Qy  2325 CGTGTTCCCGGCCGAGCCAGCACCAAGGGGGGCGCGTGTGGGACTGCCGTCCTGCTTCGTCCGGCCT 2384

Db  1629 CGTGTTCCCGGCCGAGCCAGCACCAAGGGGGGCGC 1688
         |||||||||||||||||||||||||||||||||||
Qy  2385 GCTGCTGGAATTACAGGAAGTGCTCCCCGCCCANCTGGCGTGCCTTCATCAACAAGTTTGTGTCA 2444

Db  1689 GCTGCTGGAATTACAGGAAGTTCCTCCCCGCCCAN 1748
         |||||||||||||||||||||||||||||||||||
Qy  2445 GTTCATCCATAAGTACAATTACCTTACAANGCCCAGCCCATCTCCTTCATCAACAAGTTTGTGTCA 2504

Db  1749 GTTCATCCATAAGTACAATTACCTTACAANGCCCA 1808
         |||||||||||||||||||||||||||||||||||
Qy  2505 CGCCGACCCGCTCCACGACCTGTCCTTCGACAACAGTGACCTGGTGAGCTGAAATCCCT 2564

Db  1809 CGCCGACCCGCTCCACGACCTGTCCTTCGACAACA 1868
         |||||||||||||||||||||||||||||||||||
Qy  2565 CCTTGCAGGGCTCAGCCTGCCCAGCAGGGACGAGGCCTGGACGAAGA 2624

Db  1869 CCTTGCAGGGCTCAGCCTGCCCAGCAGGGACGAGG 1928
         |||||||||||||||||||||||||||||||||||
Qy  2625 GGGCGAGGAGGAGAGCTCAGCCGGCTCCTTGGCTCAGCGTCTCCCTGTTCACCCC 2684

Db  1929 GGGCGAGGAGGAGAGCTCAGCCGGCTCCTTGGCTC 1988
         |||||||||||||||||||||||||||||||||||
Qy  2685 TCTGACCGGCGGCGCCGAGAGGGCCCCCTTACAGATGAAACGGCTGTGGA 2744

Db  1989 TCTGACCGGCGGCGCCGAGAGGGCCCCCTTACAGA 2048
         |||||||||||||||||||||||||||||||||||
Qy  2745 GGATCTGCTCGGAGGTTCTGAGTGACATAGACGAAGATGTCCGGGGAGACCCGAGANCCT 2804

Db  2049 GGATCTGCTCGGAGGTTCTGAGTGACATAGACGAA 2108
         |||||||||||||||||||||||||||||||||||
Qy  2805 CCTTGCCACGTTCACGGCTGATGAGCTCGGCCAGAGGAGTGGTTGCCGGCAA 2864

Db  2109 GAGCTTCTTCTGACCAACCTGCACGGCTGATGAGC 2168
         |||||||||||||||||||||||||||||||||||
Qy  2865 CCTCGCCTTCAGCCTTGGCCCTTGCCCTTCAGAGAACAGCATTGCAGCCGCGCTTT 2924

Db  2169 CCTCGCCTTCAGCCTTGGCCCTTGCCCTTCAGAGA 2228
         |||||||||||||||||||||||||||||||||||
Qy  2925 CCTTGCCACGTTCATGATTGTATCTGCCCTGGGCCAGCCAGCAATTGACGCAGACCGAGATCCTT 2984

Db  2229 CCTTGCCACGTTCATGATTGTATCTGCCCTGGGCC 2288
         |||||||||||||||||||||||||||||||||||
Qy  2985 CCGGAACCTTGCCTCAGTGATGAGGTGGTTGCAGAGGGCCCCTT 3044

Db  2289 CCGGAACCTTGCCTCAGTGATGAGGTGGTTGCAGA 2348
         |||||||||||||||||||||||||||||||||||
Qy  3045 GGCCTTCCTTGGTTGGTGGGCATTGACGAGCAGCTTCTCCGAGGCCCT 3104

Db  2349 GGCCTTCCTTGGTTGGTGGGCATTGACGAGCAGCT 3104
         |||||||||||||||||||||||||||||||||||
Qy  3105 GAGGATCCTTGGCATTATGGAGCCCGGTGATCGTGAGCCGCACCCCCTTCCGAGGCC 3164

Db  3105 GAGGATCCTTGGCATTATGGAGCCCGGTGATCGTG 3164
```

RESULT 5
```
LOCUS       BC018777                2418 bp    mRNA    linear   PRI 11-DEC-2001
DEFINITION  Homo sapiens, clone IMAGE:4869992, mRNA, partial cds.
ACCESSION   BC018777
VERSION     BC018777.1  GI:17511852
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2418)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 40 Row: k Column: 14
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1..2418
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4869992"
                     /tissue_type="Skin, melanotic melanoma, high MDR."
                     /clone_lib="NIH_MGC_49"
                     /lab_host="DH10B-R"
```

```
CDS    /note="Vector: pOTB7"
       <1..2097
       /codon_start=1
       /product="Unknown (protein for IMAGE:4869992)"
       /protein_id="AAH18777.1"
       /db_xref="GI:17511853"
       /translation="RLSDVRGLLRLAEAIAFRQDLEVVSTVRAVIALLRSGEQCSV
       EPDLISKVLQGLIEVRSPHLEHLLTAFFSAYADAAASPPPACKPVVVSSLILLQEERL
       AGKPGADGGSLEAVRLGPSSLLLVDWLRMLDPEVVSSCPDLQLRILFSRRKKGQAQ
       VPSRPYLLTLLFTHQSSWPTLHQCIRIVLLGKSRQRFDPSASLDLWACIHVPRIWQG
       RDQHFQKRREELVLRVQPELISLVELILAEAETRSQGDTAACSLIQARLPILLSC
       CCGDDESVRKVTEHLSGCIQQWGDSVLGRRCRDLLILQIYLQRPELRVPPEVLLHSEG
       AASSVCKIDGLIHRFILIADYSDSRALENRGADASMACRIAVAHPLLLLRHLPMI
       AALLHGRIHLNPQEFRQQNHLSCFLEVLGLLELLQPHVFRSEHQAILWDCILSFIRLI
       LNTRKSSRHLAAAFINKFVQFIHKYIYIVNAPAAISFLQKHADPLHDLSFDNSDLIVMLKS
       LIAGLSLPSRDRDIRDGLDEGEEESSAGSLPIVSVSLFTLPLTRAAEMAPIYMKRLSRGQ
       TVEDLLEVLSDIDEMSRRPEILSFFSTNLQRLMSSAPECCRNLAFSLALRSMQNSPS
       IAAAFLPTFMYCLLGSSQDFEVVQTALRNLPEYALLCQEHHAAVILHRAFLVGMYGQQMDPS
       AQISEALRILEMEAVM"

Query Match    Best Local Similarity    69.8%;    Score 2397.4;    DB 9;    Length 2418;
Matches 2398; Conservative    100.0%;    Pred. No. 0;    Mismatches    1;    Indels    0;    Gaps    0;

ORIGIN

BASE COUNT    411 a    811 c    776 g    420 t

Qy   1039  AGGCTCAGTGAGTGGTGCAGCTCAGCTCCACCGTCCGTCGCCTGCGCCTGGCCGGGGGCCTTCCGCTCCGGCCGGGCCCCTTCCGTCAG 1098
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1     AGGCTCAGTGAGTGTGCAGCCAGAGGGGGGGGGCTTCCACCTGGCCCTCCGCGCCGCCGCCGAGCGCCCCTTCCGTCAG 60

Qy   1099  GACCTGGAGGTGGTCAGCTCCACCGTCCGTCCGTGCCGTCATCGCCGCCACCTGAGGTCTGGGGAG 1158
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61    GACCTGGAGGTGGTCAGCTCCACCGTCCGTCGCTGAGGTCTGGGGAG 120

Qy   1159  CAGTGCAGGTGGAGCCGGACCTGAATCAGCAAAGTCCTCCAGGGGAGGTCGATGAGGTGAGG 1218
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121   CAGTGCAGGCTGGAGCCGGGACCTGATCAGCAAAGTCCTCCAGGGGGTCGATGAGGTGAGG 180

Qy   1219  TCCCCCCACCTGGAGGAGCTGCTGCAGCTCCATTCTTCTCTGCGCCACTGCGGGGATGCCTCC 1278
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181   TCCCCCCACCTGGAGGAGCTGCTGACGCCATTCTTCTCTGCGCGATGCGGGGATGCCTCC 240

Qy   1279  CCGTTTCCAGCTGTGTAAGCCCGTTGTGGTGTTGGTGAGCTCCTGCTTGAGCAGGAGGAGGAG 1338
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241   CCGTTTCCAGCCTGTAAGCCCGTTGTGGTGTTGGTGACGCATCTTTCTCTGCTGCTGAGGAGGAG 300

Qy   1339  CCCCTGGCTGTGGGGAAGCACGGGGTGGGCGACCGTGCGACCGTGCGGCGGCTGGGG 1398
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301   CCCCTGGCTGTGGGGGAAGCACGGGGTGGGCGACCGTGCAGCGGCGGCTGGGG 360

Qy   1399  CCCTCGTCAGGGCCTCCTAGTGGACTGGGCTGGACCCCGGAGTGGGTCAGCAGC 1458
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361   CCCTCGTCAGGGCCTCCTAGTGGACTGGGAAATGCTGGACCCCGGAGTGGGTCAGCAGC 420

Qy   1459  TGCCCCGGACCTGCAGCTGCAGCTGCTCTCTTCCCCGGAGAGGAGGAGGGCCAG 1518
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421   TGCCCCGGACCTGCAGCTCAGCTGCTCTCTTCCCCGGAGAGGGCAAAGGTCAGGGCCCAG 480

Qy   1519  GTGCCCTCGTTCCGTTCCGTCCCTGACCCTCTTCACGCATCAGTCCACTGGCCCACA 1578
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481   GTGCCCTTCGTTCCGTCCCTACCCTCCTCAGCGACATCAGTCCACTGGCCCCACA 540

Qy   1579  CTGCACCAGTGCATCCGAGTCCTGTGGGCAAAGAGCCGGGAACAGAGGTTCGACCCCTCT 1638
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541   CTGCACCAGTGCATCCGAGTCCTGTGGGCAAAGAGCCGGGAACAGAGGTTCGACCCCTCT 600

Qy   1639  GCCTCTCTGGACTTCCTCTGGGGGCCTCTGCAGCGGGGGGCCGGGAC 1698
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   601   GCCTCTCTGGACTTCCTCTGGGGCCTCCATCCATGTGCCAGGGGGCCGGGAC 660

Qy   1699  CAGCGCACCCCGCAGAACGGGGGAGGAGAGGAGCTGGTGCTGCTGCTGTCGGGGTTCGCGGGGTCCCGGGAGCTC 1758
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661   CAGCGCACCCCGCAGAAGCGGGGGAGGAGGAGCTGGTGCTGCTGGGGTTCCAGGGGCCCGGGAGCTC 720

Qy   1759  ATCAGCCTGGTGGAGCTGATCCTGGCCGAGCCGGGGCGGGGGGAGGAGGCCGAGGCCGAGGAGGACACA 1818
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721   ATCAGCCTGGTGGAGCTGATCCTGGCCGAGCCGGGGCTGCTCAGCTGTGTGGGGACA 780

Qy   1819  GCCGCCTCAGCGCCATCAGCAGGGCGCAGACCTGGAGGTGCTGGGGGCGGGGGGGCGGGGGACA 1878
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781   GCCGCCGCCAGCGCCATCAGCAGGGCGCCGGGCGGGGCGGGGGGGGGGGGGGGGGAC 840

Qy   1879  GATGAGAGTGTCAGGAAAGTGGAGGAGGACACCTGCGATCCAGCAGTGGGGGGAC 1938
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841   GATGAGAGTGTCAGGAAAGTGACGGAGGAGCACCTGCAGCGGGGGAC 900

Qy   1939  AGGGTGCTGGGGAAGGCGGCTGGGCCGGGAGACCTTCCTCCTCGCAGCTCACCTACGACGGCGGGGGAG 1998
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901   AGGGTGCTGGGGAAGGCGGCTGGGCCGGGAGACCTTCCTCCTCGCAGCTCACCTACGACGGCGGGGGAG 960

Qy   1999  CTGCGGGGTGGCGGCGGCTGCCCTGACGGTCCTTACTCGCCAGCCGAAGGGGCTGCCAGCAGCACGGTC 2058
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961   CTGCGGGGTGGCGGCCGGCTGCCCTGAGGTGTTACTCGCCAGCCGAAGGGGTCGCGAGCAGCAGCACGGTC 1020

Qy   2059  TGCAAGCTGGACGGAGGACTCATCACCGCTTCATCACGGCTTCCTTGGGGACACCAGCCAGGACTCC 2118
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  TGCAAGCTGGACGGAGGACTCATCATCACCGGCTTCATCACGGTCTCCTTGGGGACACCAGCCAGGACTCC 1080

Qy   2119  CGGGGCGTTGGAGAACCAGGAGGCGCAGATGGCGGCGCGCATGGCCGTGCGCGAGGCCTGGGGCGGG 2178
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  CGGGGCGTTGGAGAACCAGGAGGCGCAGATGGCCAGCAGGCGGGGAGGCGTGCGCGAGGCCTGGGGC 1140

Qy   2179  CACCCCGCTGCTGCTGCTTCAGGCACCTGCCCATGAATCGGGCCTCCTCCACGGGCCACC 2238
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  CACCCCGCTGCTGGTTGCTCAGGCACCTGCCCATGAATCGGGCCCATGATGGGCCACC 1200

Qy   2239  CACCTCCAACTTCCAGGAGTTCCGGCAGCAGGAACCACCTGAGCTCGTTCCTTCCTGCACTGCTG 2298
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  CACCTCCAACTTCCACAGGAGTTCCGGCAGCAGGAACCACCTGAGCTCGTTCCTTCCTGCACTGCTG 1260

Qy   2299  GGCCTGCTGGAGCTGCTGCAGCCGCACTGTGTTCCGGCAGCCACGAGGGCGGCCCTGTGG 2358
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  GGCCGCTGGAGCTGCTGCAGCCGCACTGTGTCCGGCAGCACCTGAGCTCGCCTTCCTGCACCGTGTGG 1320

Qy   2359  GACTGCCTTCTGTGTCCTTCATCCCGCCTGCTGGAATTACAGGAAGTCCTCCCGGGCCATCTG 2418
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  GACTGCCTTCTGTGTCCTTCATCCCGCCTGCTGAATTACAGGAAGTCCTCCCGGGCCATCTG 1380

Qy   2419  GCTGCCTTCATCAAACAAGAGTTGTCGCAGTTCATCCATAAAGTAACATTACCTACAAATGCCCCA 2478
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  GCTGCCTTCATCAACAACAAGAGTTTGTCGCAGTTCATCCATAAAGTACATTACCTACAAATGCCCCA 1440

Qy   2479  GCAGCCAATCTCCTTCCCTGCAGAACCAGCCGACCCGGCTCCAGCAGCACCTGTCCTTGGACAAC 2538
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  GCAGCCAATCTCCTTCCCTGCAGAACCAGCCGACCCGGCTCCAGCAGCACCTGTCCTTGGACAAC 1500

Qy   2539  AGTGACCTGGTGATGCTGAAATCCCTGCTTGCAGGGCTCAGCCTGCCCAGCCGGGACCGAC 2598
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501  AGTGACCTGGTGATGCTGAAATCCCTGCTTGCAGGGCTCAGCCTGCCCAGCCGGGACCGAC 1560

Qy   2599  AGGACCGAGGACCTGGACGAAGGAGGAGAGGAGGAGTCAGCTGGCTCCTTGCC 2658
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  AGGACCGAGGACCTGGACGAAGGAGGAGAGGAGGAGTCAGCTGGCTCCTTGCC 1620

Qy   2659  CTGGTCACGGTCCTGTTCACCCTGCCGCTGACCCGGGCCGCGGAGTGCTTCCTTGCC 2718
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621  CTGGTCACGGTCCTCCTGTTCACCCTGCCGCTGACCCGGGCCGAGATGGCCCCTACATGAAA 1680

Qy   2719  CGGCGTTCCCGGGCCCAAACGGTGGAGAATCTGCGTCGAGAGTTCTGAGTGACAATAGACGAG 2778
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681  CGGCGTTTCCCGGGCCCAAACGGTGGAGAATCTGAGTGACAATAGACGAG 1740

Qy   2779  ATGTCCCGGCCGGAGACCCGAGAtCCAGGTCTTCTCTGACCAACCTGCAGCGCGGGCTGATG 2838
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741  ATGTCCCGGCCGCGGAGACCCGAGATCCAGGTCTTCTCTGACCAACCTGCAGCGCGGGCTGATG 1800
```

```
RESULT 6
LOCUS       HSM800197               2241 bp    mRNA    linear   PRI 18-FEB-2000
DEFINITION  Homo sapiens mRNA; cDNA DKFZp586J0619 (from clone DKFZp586J0619);
            partial cds.
ACCESSION   AL050110
VERSION     AL050110.1  GI:4884139
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2241)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-1999) MIPS, Am Klopferspitz 18a, D-82152
            Martinsried, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp586J0619) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1..2241
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp586J0619"
                     /tissue_type="uterus"
     gene            1..1924
                     /gene="DKFZp586J0619"
     CDS             <1..1924
                     /gene="DKFZp586J0619"
                     /note="unknown"
                     /codon_start=2
                     /product="hypothetical protein"
                     /protein_id="CAB43278.1"
                     /db_xref="GI:4884140"
                     /db_xref="SPTREMBL:Q9Y3W8"
                     /translation="TSSRSRAELLILTAFFSATADAASFPPACKPVVVSSLLLQEEEPL
                     AGGKPGADGGSIEAVRLGPSSGLLVDWLRMLDFPEVVSSCPDLQLRLIFSRRGKGQAQ
                     VPSFRPYILILFTHQSSWPTLHQCIRVLLGKSRSQRDFDSASIDFIWACIHYPRIWQG
                     RDQRTPQKRREELVLRVGPPELISIVELILLAEAETRSQDGDTRACSLIQARLPLLISC
                     CCQDDESVRKVTEHLSGGIQOWRGDSVLGRRCRDILLQLYLQRPELRVPVPEYLLHSEG
                     AASSSVCKLLDGLIHRFITLLADYISDRALENRGADASNACRKLIAYAHPLLLLRHLPMI
                     AALLHGRTHLNFQEFRQQNHLSCFLHIVLLELLQPHVFRSEHQGALWDCLLSFIRLL
                     LNYRKSSRHLAAFINKFVQFIHKYITYNAPAAISFLQKHIADPLHDLSPDNSDLVMLKS
                     LLAGLSLPSRDDRTDRGLDEEGEESSAGSLPIVSVSLFTPLTAAEMAPYMKRLLSRGQ
                     TVEDLLEVLSDIDEMSRRPFEILSFFSTNLQRLMSSAEEECRNLAFSILAIRSMQNSPS
                     IAAAFLPTFMYCLGSQDPEVVQTALRNLPEYALLCQEHAAVLLHRAFIVGMGQMDPS
                     AQISEALRILHMEAVM"
     polyA_site      2224

BASE COUNT      383 a    754 c    712 g    392 t
ORIGIN
```

```
Query Match     64.1%;   Score 2203.8;   DB 9;   Length 2241;
Best Local Similarity   99.8%;
Matches 2205;   Conservative  0;   Mismatches  2;   Indels  0;   Gaps  0;
                                   Pred. No. 1.6e-304;
```

```
Qy     80 TGTAAGCCCGTTGTGGTTGGTTGAGCTCCCTGCTGCTGCAGGAGGAGGAGCCCCTGGCTGGG  139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1291 TGTAAGCCCGTTGTGGTTGGTTGAGCTCCCTGCTGCTGCAGGAGGAGGAGCCCCTGGCTGGG 1350

Qy     20 GCGGGACCTGCTGACTGCGATTCTTCTTCTCGCCACTGCGGATGCTGCCTCCCCGTTTCCAGCC  79
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1231 GAGGAGCTGCCTGACTGCGATTCTTCTTCTCGCCACTGCGGATGCTGCCTCCCCGTTTCCAGCC 1290

Qy    140 GGGAAGCCGGGTGCGGACGGTGGCCAGCCCTGGAGGCCGTGCGGCCCTCGGTCAGGC  199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1351 GGGAAGCCGGGTGCGGACGGTGGCCAGCCCTGGAGGCCGTGCGGCCCTCGGTCAGGC 1410

Qy    200 CTCCTAGTGGACTGGCGTGGAGAAATGCTGGACCCCGAGGTGGTCAGCAGCTGCCCCGACCTG  259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1471 CTCCTAGTGGACTGGCGTGGAGAAATGCTGGACCCCGAGGTGGTCAGCAGCTGCCCCGACCTG 1470

Qy    260 CAGCTCAGGCGCTGCTCTTTCTCCCCGGAGGAAGGCGCAAAGGTCAGGCCCCAGGTGCCCTTCGTTC  319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1411 CAGCTCAGGCGCTGCTCTTTCTCCCCGGAGGAAGGCGCAAAGGTCAGGCCCCAGGTGCCCTTCGTTC 1530

Qy    320 CGTCCCTACCTCCTCAGCCACCTGCGGCCCATCAGTCCAGCTGCCCACACTGCCACCAGTGC  379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1531 CGTCCCTACCTCCTCAGCCACCTGCGGCCCATCAGTCCAGCTGCCCACACTGCCACCAGTGC 1590

Qy    380 ATCCGAGTTCCTGGTCGGACGCCGGGACACAGAGAGTTCGACCCCTGCCTCGTTC  439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1591 ATCCGAGTTCCTGGTCGGACGCCGGGACACAGAGAGTTCGACCCCTGCCTCGTTC 1650

Qy    440 TTCCTCTGGGCCCGCCATCCAATGTTCCTCGCATCTGGCAGGGGCGGGGACCAGGCGACCACCCG  499
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1651 TTCCTCTGGGCCCGCCATCCAATGTTCCTCGCATCTGGCAGGGGCGGGGACCAGGCGACCACCCG 1710

Qy    500 CAGAAGCGGGGGAGGAGCTGGTCGTCGGCCCGGGCCCGGAGCTCATCAGCCTGGTG  559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1711 CAGAAGCGGGGGAGGAGCTGGTCGTCGGCCCGGGCCCGGAGCTCATCAGCCTGGTG 1770

Qy    560 GAGCTGGATCCTGGCCGGAGCGGGACGGGGGACACAGCCGGCCTGCAGC  619
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1771 GAGCTGGATCCTGGCCGGAGCGGGACGGGGGACACAGCCGGCCTGCAGC 1830
```

```
Qy  1831 CTCANCCAGGCAGCCCGGCTGCCCGGCTGCTCAGCTGTCTGCTGTGTGGGGACGATGAGAGTGTC 1890
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   620 CTCANCCAGGCAGCCCGGCTGCCCGGCTGCTCAGCTGTCTGCTGTGTGGGGACGATGAGAGTGTC  679

Qy  1891 AGGAAGGTGACGGAGCGAGCACGTGCTGGGA 1950
         ||||||||||||||||||||||||||||||
Db   680 AGGAAGGTGACGGAGCGAGCACGTGCTGGGC  739

Qy  1951 AGGGCGCCTGCCGAGACCTTTCTGCGAC 2010
         ||||||||||||||||||||||||||
Db   740 AGGGCGCCTGCCGAGACCTTTCTGCGACGTGCGGGTGCCC  799

Qy  2011 GTGCCTGAGGTCTCTACTGCGACACGAGGGAGAGGGCGTC 2070
         ||||||||||||||||||||||||||||||||||
Db   800 GTGCCTGAGGTCTACTGCGACACGAGGGAGCTGGGAC  859

Qy  2071 GGACTCATCCACCGCTTCATCACGCTCCTTGTTCGGAC 2130
         ||||||||||||||||||||||||||||||||
Db   860 GGACTCATCCACCGCTTCATCACGCTCCTTGTTCGGAC  919

Qy  2131 AACCGAGGGGGGATGCCAGCATGGCCGTGGGGTTGGGG 2190
         ||||||||||||||||||||||||||||||||||
Db   920 AACCGAGGGGGGATGCCAGCATGGCCGTGGGGTTGGGG  979

Qy  2191 CTGCTCAGGCACCTGCCCATGATCGGGGCTCCTG 2250
         ||||||||||||||||||||||||||||||
Db   980 CTGCTCAGGCACCTGCCCATGATCGGGGCTCCTGCCACCTC  1039

Qy  2251 CAGGAGTTCCGGCAGCAGGAACCACCTCCTTGTTCCTG 2310
         ||||||||||||||||||||||||||||||||
Db  1040 CAGGAGTTCCGGCAGCAGGAACCACCTCCTTGTTCCTG  1099

Qy  2311 CTGCTGCCAGCGCGCACGTGTTCCGCAGGCAGG 2370
         ||||||||||||||||||||||||||||||
Db  1100 CTGCTGCCAGCGCGCACGTGTTCCGCAGGCAGG  1159

Qy  2371 TCCTTCATCCGCTCGCTGCTGAATTACAGGAGG 2430
         ||||||||||||||||||||||||||||||
Db  1160 TCCTTCATCCGCTCGCTGCTGAATTACAGGAGG  1219

Qy  2431 AACAAGTTTGTGCAGTTCATCCATAAGTACA 2490
         ||||||||||||||||||||||||||||
Db  1220 AACAAGTTTGTGCAGTTCATCCATAAGTACA  1279

Qy  2491 TTCCTGCCAGGAGCACGCCGACCGGTCCCACG 2550
         ||||||||||||||||||||||||||||||
Db  1280 TTCCTGCCAGGAGCACGCCGACCGGTCCCACG  1339

Qy  2551 ATGCTGAAATCCCTCCTTGCAGGGTCAGCC 2610
         ||||||||||||||||||||||||||||
Db  1340 ATGCTGAAATCCCTCCTTGCAGGGTCAGCC  1399

Qy  2611 GGCCTGGACGAGAAGCACCGCGAGGAGAGC 2670
         ||||||||||||||||||||||||||||
Db  1400 GGCCTGGACGAGAAGCACCGCGAGGAGAGC  1459

Qy  2671 TCCCTGTTCACCCCTCTGACCGGCGGCGGC 2730
         ||||||||||||||||||||||||||||
Db  1460 TCCCTGTTCACCCCTCTGACCGGCGGCGGC  1519

Qy  2731 GGCCAAACGGTGGAGGATCTGCTGGAGGTTG 2790
         ||||||||||||||||||||||||||||
Db  1520 GGCCAAACGGTGGAGGATCTGCTGGAGGTTG  1579

Qy  2791 AGACCCAGAGTCCTGAGCTTCTTCTCGACCAAC 2850
         ||||||||||||||||||||||||||||
Db  1580 AGACCCAGAGTCCTGAGCTTCTTCTCGACCAAC  1639

Qy  2851 GAGNGTTGCCCGAACCTGGCCTTCAGCCTGG 2910
         ||||||||||||||||||||||||||||
Db  1640 GAGNGTTGCCCGAACCTGGCCTTCAGCCTGG  1699

Qy  2911 ATTGCAGCGCGCTTTCCTGCCCAGTTGTCATG 2970
         ||||||||||||||||||||||||||||
Db  1700 ATTGCAGCGCGCTTTCCTGCCCAGTTGTCATG  ...
```

```
Db  1760 GTGCAGACGGCGCCTCCGGAACCTGCCCACGTTCAT 1819
         ||||||||||||||||||||||||||||||||

Qy  2971 GTGCGACGGCGCCTCCGGAACCTGCCCACGTTCAT 3030
         ||||||||||||||||||||||||||||||||
Db  1820 GTGCTGCTTCCACCGGCCCTTCCTGGTGGCAC  1879

Qy  3031 GTGCGTGCTCCACCGGCCCTTCCTGGTGGCAC 3090
         ||||||||||||||||||||||||||||
Db  1880 ATCTCCGAGGCCCTGCAGGAATCCTGACCG  1939

Qy  3091 ATCTCCGAGGCCCTGCAGGAATCCTGACCG 3150
         ||||||||||||||||||||||||||||
Db  1940 CCCCCTCCAAGCCCCGTCCCCGGGGACCG  1999

Qy  3151 CCCCCTCCAAGCCCCGTCCCCGGGGACCG 3210
         ||||||||||||||||||||||||||||
Db  2000 CGTGGGGCGTTGCTGGTTCTGTCCGCCGAGAGG  2059

Qy  3211 CGTGGGGCGTTGCTGGTTCTGTCCGCCGAGAGG 3270
         ||||||||||||||||||||||||||||
Db  2060 CCAGGAGCAAATACTCCGACCCCTGGGGTGGCCG  2119

Qy  3271 CCAGGAGCAAATACTCCGACCCCTGGGGTGGCCG 3330
         ||||||||||||||||||||||||||||
Db  2120 TCCAGCAAGCCCTCATTCACCTTCTGGGCC  2179

Qy  3331 TCCAGCAAGCCCTCATTCACCTTCTGGGCC 3390
         ||||||||||||||||||||||||||||
Db  2180 CGGGCATGGGCCTGGGCTGGTTTTGAATGAAACAGACCTGAACTGTCAA 2226

Qy  3391 CGGGCATGGGCCTGGGCTGGTTTTGAATGAAACAGACCTGAACTGTCAA 3437
```

RESULT 7

```
LOCUS       BC010333                2730 bp    mRNA    linear   ROD 07-AUG-2002
DEFINITION  Mus musculus, clone IMAGE:3601115, mRNA, partial cds.
ACCESSION   BC010333
VERSION     BC010333.1  GI:16307573
KEYWORDS    
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2730)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (05-JUL-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
COMMENT     NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
REMARK      Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 16 Row: n Column: 18
            This clone was selected for full length sequencing because it
```

passed the following selection criteria: Hexamer frequency ORF
analysis.

```
FEATURES          Location/Qualifiers
     source       1..2730
                  /organism="Mus musculus"
                  /db_xref="taxon:10090"
                  /map="FVB/N"
                  /clone="IMAGE:3601115"
                  /tissue_type="Mammary tumor. Metallothionien-TGF alpha
                  model. 10 month old virgin mouse. Taken by biopsy."
                  /clone_lib="NCI_CGAP_Mam1"
                  /lab_host="DH10B"
                  /note="vector: pCMV-SPORT6"
     CDS          <1..2458
                  /codon_start=2
                  /product="Unknown (protein for IMAGE:3601115)."
                  /protein_id="AAH10033.1"
                  /db_xref="GI:16307574"
                  /translation="IALALQQALGQELARVRQGNPEVPGITVRLLQAMTTLLSSPEGG
                  TLALAMHHSRFLSCPLMRQLIQYQRAVPQDTGFSSSLFLKVLMQLLOWLDSPAVEDGPL
                  QAQLKLPATRYSANHRISDVRSGLLHIADALSFEHGDLEVANSTARAVIATLRSGEKCP
                  VEPELLSKVLRGLIEVRSPHLEEILLTALFSAATTETSCPSPASGFIVVVSSLLLQEKEH
                  LLGPSKQEVEGASTEAWRLGPASGLLVDWLWETLDPEVVCSCPDLQWKLLFSRRKGKGH
                  ISAQVLSFRPYILALLTHQASWSTLHRCIRVILGKSREQRLDPSASLDFLWACIHVPR
                  IWQGRDQRTPQKRREELVIHVQGPELISVELILSEAFTRSQQDSAARTLIQTRLPL
                  LLSCCRSNDESIGKVTEHLTSCIOOMGDSVLGQRCRDLLLQLYLQRPEVRVPVPEVIL
                  QSEGATSSSICKLDGLVHRFITILLADTSDSRSSESRVADANMACRKLAVAHPVLLIRH
                  LPMIALLHQRTLHNFQEFRQQNHLAFFLHVLGIELLQPRVFQSEHQGALWDCLRSF
                  IRLLINYRKSSRHLAPFISKFVQFIHEKYVGCSAPAAYAFLQKHAEPLHDLSPDNSDLV
                  MLKSLLAGLSLPSRDGRIDQGLDEEGEDREBRSAGSILPIVSVSLSTPLIVADVAPHKKRL
                  SRQRAVEDVLETLLSDIIDEMSRRRPEVLGRFSINLQRLMSSAEESCRNLAFSLALRSIQ
                  NNPSTAADFLPTFMYCLGSRDFPEVVQTALRNLPEYTLLQQHAAVILHRAFLVGYVGQ
                  IDFSAQISEALKILHMRAVM"
```

```
BASE COUNT    541 a    877 c    786 g    526 t
ORIGIN

Query Match     50.2%;    Score 1726;    DB 10;    Length 2730;
Best Local Similarity    82.0%;    Pred. No. 1.7e-236;
Matches 2016;    Conservative   0;    Mismatches 435;    Indels   9;    Gaps   2;
```

```
Qy    687  CGTGGCCTGCCCTGCCCTCCAGCAGGCCCGCCCTGGGCCCGCCGAGCTGGCCGTCCAGGCAG  746
           | |||||| ||  || |||| ||| || ||||||||||| || |  | ||||||||||||||
Db      1  CCTTGGCCCTGGCACTCGCAGCAAGCCCTGGGCGCCAGAGTTGGCCGGTTGGTCCGGGGAA  60

Qy    747  CCCCGGAGGTGCCGGGGCATCAGCAGCCCTTCGTGTCCTCGCGCTCGCTCGCTCACTC  806
           ||||| |||||||| || |||| |||||| || ||| ||||||||||||||| ||||||
Db     61  CCCCGGAGGTGCCAGGCGATCACAGTCCCGCCCTTCTGCAGGCCATGACCACCCTGCTGAGCTC  120

Qy    807  CCCACACGGCGGGTGCCCTGGTGTGAGTCCATCGCACCGTAGCCACTTCCTGCCCTGCCCGCT  866
           || ||  || ||| ||||| || ||||||||| ||||| || |||| ||||||||||||||||
Db    121  CCCGCACGGTGGCCACCCTTGCTCTGCTATGCATCACAGCCACTTCCTTTCTGCCCCACT  180

Qy    867  GCTGCGCCAGCTCTCCAGTACCAGGCGCTGTGTGCCCACAGGACCACGGCTTCTCCTCGCT  926
           || | |||||||||| |||| |||||||| || | ||| ||||||||||||||||||||||
Db    181  GATGCGGCCAGCTCTACCAGTACCAGGCGTGCTGTCCCGGAGCACACTGCCGCTTCTCCTCGCT  240

Qy    927  CTTCCTCGAAGGTGGCTCCCTGCAGATCGCAGATGGCTGCAGTGGCTGGCGTGGAGGGCG  986
           |||| |||||||||||||||||||||||||||| ||||||||||||||| ||||||||||
Db    241  CTTCCTTAAAGTGTCATCGCAGATCCTGCAGTGGCTGGATAGCCCCTGCTGTGAGGGATGG  300

Qy    987  GCCCCTTCGCGGGGCCACAGCTCAGGAGGTGTTGCCAGCCCAGGGCGGCCAGGCTCAG  1046
           |||||| | |||  ||||| ||| |||||||||||||| |||||||||| ||||||||||
Db    301  GCCCTTGCAGGCCCAACTCAAGTTATTTGCCTACTGCGCTTATTCAGCAAGGCACAGGATCAG  360

Qy   1047  TGATGGTGCGAGGGGGGGGGGGCTCGCGCGTCGGCGGCGCCAGGGCCCGCCTTGCAGGACCTGGA  1106
           ||||||||||| |||| || |||| || |||||||||||| ||||||||||||||||||
Db    361  CGACGTGCGTAGTGGGGCCTTCCTGCACTTGGCGTGACGCCTGAGCGTTCACGGGTGAGCTGGA  420

Qy   1107  GGTGGGTCAGTCACCGTCCACCGGTCACCGTGCCGTCATCGCCACCCTGAGGTCTGGGGACAGTGCAG  1166
           |||| || || || |||| |||||| ||||||||||| |||||||||||||||| ||||
Db    421  GGTGGCCAACTCACCGCGCGGGGCGTCGATCAGCCCTCAGGTCCGGGGAGAAGTGTCC  480

Qy   1167  CGTGGAGCCGGACCTGATCAGCAAAGTCCTCCAGGGGCTGCAATCGCAGTCGAGGTCCCCCCA  1226
           |||| || ||||| || ||||| ||||||||||||  | || | ||| ||||| |||||||||
Db    421  GGTGGCCAACTCACCGCGCGGGGCGTCGATCAGCCCTCAGGTCCGGGGAGAAGTGTCC  480
```

```
Db    481  TGTGGAGCCTGCAGCTCATCAGCAAAGTCCTCCGGGGGCCTTCATCGGAGGTGGCTCCCGCA  540
           |||||||| ||| ||||||| | ||||| | || |||||||| |||| |||||| ||||||
Qy   1227  CCTTGGAGGAGCTGCGCTGACTGGCATTTCTTCTCTGGCCACTGGGATGCTGCCCCGTTTCC  1286
           |||||||| |||| || || ||  || |  | | || || ||||||||||||||| ||||||
Db    541  CTTTGGAGAGTTGCTGACAGCGCTCTTGCTCAGCGCCACCACAGAGAACCTTCGTGCCCGACCC  600

Qy   1287  AGCCTGTTAAGCCCCGTTTGTGGTGGGTGACCTCCCTGCTGCTCGCAGG---AGGAGGAGCCCCT  1343
           |||||| ||| ||||| || || |||| || || |||||||| |||||   ||||||||||||
Db    601  AGCCTTCGGGCCCCCATTGTGGTGGGTGAGCCTCTTGCTGCTACAGGAGAAGGAAGAGCTCCT  660

Qy   1344  GGCTGGGGGGGAAGGCGGGTTGCCGAGGGCGGACGGTGGCAGCGGCCGTGCGGGCTGGGGGCCCTC  1403
           ||||| || || ||| || || || |||| |||| |||| || || |||||||||| |||||
Db    661  GGGGCCCCAGCAAGCAGCAGGAGGTGCCAGCACGGAAGCCATGCGCGCCCTGGGGACCAGC  720

Qy   1404  GTCAGGCCCTCCTTAGTTGTGGGACTGGGCTGGAAAATGCTGGACCCCGGAGGTGGGTCAGCAGCTGCCC  1463
           | ||||| ||| ||| | || || ||||| |||||| | || || |||||| || | |||||| ||
Db    721  TTCTGGGCGTGCTTGTTGGGACTGGAAGCCTGGAGAGAGGCCGGAGACCCCTGTGTAGTTGCCC  780

Qy   1464  CGACCTGCAGCTCAGGCTGCCTCTTCTCCCCGGAGGGAAGGCAAAGGTCA-----GGCCCA  1517
           ||||| ||| || ||||||| | |||||||| ||||| | |||||| |||     |||||
Db    781  TGACCTACAGTGGAAGCTGCTCTTCTTCTCACGGAAAAGGCCAAATGCCCACATAAGTGCCCA  840

Qy   1518  GGTGCCCTCGTTCCGGTCCCTACCTCCTTCACCGCATCAGTCAGGTGGCCCAC  1577
           ||||||| || || || || | |||| ||||||||| |||| || ||||||
Db    841  GGTGCCTCTCTTCCGCCCCTCACCTCCTGCCCTTCAGTGCGTCCCGTTAGCTGCCAC  900

Qy   1578  ACTGCACCCAGTGCATCCAGAGTCCCTGGTGGCAAGAGGCCGAGAACAACAGAAGGTTCGACCCTC  1637
           |||||||| |||||||| |||||| || || || |||||| ||| ||| || || |||||||||
Db    901  GCTGCACCGTGCAATCCGTGCCTGCTAGGCAGGAGCCGGTGCTGGTTCCAC  960

Qy   1638  TGCCTCTTCTGGACTTTCTCTCGGGAGTCATCCATCTGGCCATCTGGCAGGGGGGA  1697
           || | ||||| | |||| || | ||||||  ||||||||  | |||||||||||||
Db    961  AGCATCCCTGGACTTCCTTCGGGCCTTGTAATCCATGTCGGGCCTGGCCAGGGCCAGGA  1020

Qy   1698  CCAGCGCACCCCCCGAGAAGCGGCGGGGAGGAGGAGGAGTCGGGTGTCGGGTCCAGGGCGGACT  1757
           ||||||||||||| || |||||||| | || ||||| | | |||||||| |||||||||||
Db   1021  CCAGCGCCACTCCACAGAAGGCGGGGGGAGGAGGAGTTGGTGCTCCACGGTCAGCAGGGGGGA  1080

Qy   1758  CATCAGCCCTGGTGGAGCTGATCCTGGCCGAGGGGGAGGACGGAGACGGGAGGGGACAC  1817
           |||||||| || || ||| | || |||||||| |||||||||| | || |||||||||
Db   1081  GCTCACCCTGTGGAGCTGACGTTATCCCTGCCGGGGCCCAGGGGCCCAGAGCAGGGGACAG  1140

Qy   1818  AGCCGGCCTGCCAGCCTCATCCAGCCGGGCCTGCCCTTGCGTGCTCAGGCTGGCCTTGGTGGGGC  1877
           || |||| || || ||||||| |||||||| ||||||||| |  | || |||||  |||||
Db   1141  TGCTGCACCGACCACCCTTATTCCAGACCCGGCCTGGGCTAGTGCCTTAGCTGCTGCCCGTTAGCAA  1200

Qy   1878  CGATGAGAGTGTCAGGAAGTGACGGAGCACTTGTCAGGCTGGCATCCAGTGGGAGA  1937
           ||||||||| |||| || || | ||| ||||| | ||| || ||| ||||||||
Db   1201  TGATGAGACATTGGAAAAGTGACGACCACCTGACACTCCCAGCAGCACAGCACAT  1260

Qy   1938  CAGGGTGCTGGGAAGGCGGTCGTGCCGAGACCTTCTCCTGACGCTCACTACTACGGGGGA  1997
           | |||||||||| ||||| || ||||| ||||||| | || || || || |||||||
Db   1261  CAGTGTGCTGCGGGCGCACCGCTGCCAGAGCTTGGCTGGTTGCAACTGTACCTGCAGGGGCAGA  1320

Qy   1998  GCTGCGGGTTGGCCCGGTGCCCGCGGGTCCTACTGCACGAGGGCGCAGAGGCGGCTGCCGCAGCAGCGCT  2057
           ||||| |||| ||||||| ||| || ||||| | | ||| || ||| || |||||||
Db   1321  GGTCCGGGTGCCAGGTGCCGGTCCTGCGGGTTGCCAGTGGTGCCAGCAGCAGCAT  1380

Qy   2058  CTGCAAGCTGGACGGACTCATCCACGCGGCTTCATCACGCTCCTTGGGGACACCAGCAGCT  2117
           ||||||||| ||| || || ||| || |  | || || || | ||| |||||| | | |
Db   1381  CTTGGAAGCTGGACGGCCTTCCAATGGCTTCAGCTCTGGGGGGGGA  1440

Qy   2118  CCGGGCGTTGAGAAACCGAGGGGCGGATGCCCGTCAGCCCGAGAAGCTGCGCGGTGGC  2177
           | ||  ||| | || || | || |||||||||| ||||| || |||| ||||||||
Db   1441  CCGGTGCTTCTGAGACCAGAGTGGGCTGATGCCAACATGGCCACCAGCAGACGGCGGGC  1500

Qy   2178  GCACCGGCCTGCTGCTCGCAGCCACCTGCCCAGATGTGGCCGCGCGCTTCGCCAGGGCCGCGCAC  2237
           ||||| |||||||| || ||| || |||||||||| || ||| ||| ||| || || ||||
Db   1501  CCACCCGTGCTGCTGAGGCGCACCCTCAGGGTGCCCGGGAAGCTGGCTGGGC  1560

Qy   2238  CCACCTCAACTTCCAGGAGTTCCGGCAGCAGAACCACACCTGGCGTGCTCCGTTCCTCGCAAGGTGCT  2297
           ||  ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2238  CCACCTCAACTTCCAGGAGTTCCGGCAGCAGAACCACACCTGGCGTGCTCCGTTCCTCGCAAGGTGCT  2297
```

```
RESULT 8
HSM802043
LOCUS       HSM802043              1141 bp    mRNA    linear   PRI 18-FEB-2000
DEFINITION  Homo sapiens mRNA; cDNA DKFZp434C0126 (from clone DKFZp434C0126);
            partial cds.
ACCESSION   AL137358
VERSION     AL137358.1  GI:6807876
KEYWORDS    .
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1141)
  AUTHORS   Ansorge,W., Wirkner,U., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2000) MIPS, Am Klopferspitz 18a, D-82152
            Martinsried, GERMANY
  COMMENT   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp434C0126) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1..1141
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434C0126"
                     /tissue_type="testis"
                     /clone_lib="434 (synonym: htes3). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
     gene            1..934
                     /gene="DKFZp434C0126"
     CDS             <1..934
                     /gene="DKFZp434C0126"
                     /note="DKFZp586J0619, partial"
                     /codon_start=2
                     /product="hypothetical protein"
                     /protein_id="CAB70710.1"
                     /db_xref="GI:6807877"
                     /translation="LLQPHVFRSEHQGALMDCLLSFIRILLNYRKSSRHLAAFINKFV
                     QFIHKYIYNAPAAISFLQKHADPLHDLSFPNSDLVMLKSLLAVLSLPSRDDRRDRGL
                     DEEGEEESSAGSLPLVSVSLFLTLTRAEMAPYMKRLSRQTVEDLLEVLSDIIDEMSRR
                     RPEILSFFSTNIQRLMSSAEECRNLAFCLALRSMQNSPSIAAAFLPIFMYCLGSQDF
                     EVVQFALRNLPHYALLCQEHAAVLLHRAFPGGHVRPDPGQRADLRGPEDPAYGGRDVS
                     LWQPTPLQAPARPVPGDPRGKAQEAWALLVCPRR"
     polyA_site      1126
BASE COUNT      209 a    384 c    339 g    209 t
ORIGIN
```

```
Query Match       31.4%;  Score 1079.4;  DB 9;  Length 1141;
Best Local Similarity   98.4%;
Matches 1111;  Conservative    0;  Mismatches   16;  Indels    2;  Gaps   2;
```

```
Qy       1561 TCATCTGAACTTCCAGGAGTTCCGGCAGGAGAACCACCTGGCCTTCTTCCGCAATGTGCTT 1620
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2238 ... 2298
...
```

(Sequence alignment blocks Qy/Db continue for the full 1141 bp query against the database sequence, columns paired line-by-line with match bars "|".)

```
Qy  2730 GGGGCCAAACGGTGGAGGAATGTGCTGGGAGGTTCTGAGTGAGTACATAGAGACGGAGATGTCCCGGCG 2789
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GGGGCCAAACGGTGGAGGAATGTGCTGGGAGGTTCTGAGTGAGTACATAGAGACGGAGATGTCCCGGCG  480

Qy  2790 GAGACCCAGAGATCCTGAGCTTCTTCTGGACCAACCTGCAGCGCGCGCTGATAGAGCTCGGGCCGA 2849
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGACCCAGAGATCCTGAGCTTCTTCTGGACCAACCTGCAGCGCGCGCTGATAGAGCTCGGGCCGA  540

Qy  2850 GGAGTGTTGCCGCAACCTGCCTTCAGCCTGTCCGCTCCGCTCCATGCAGCGAACAGCCCCAG 2909
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GGAGTGTTGCCGCAACCTGCCTTCAGCCTGTCCGCTCCGCTCCATGCAGCGAACAGCCCCAG  600

Qy  2910 CAATTGCAGCCCGCTTTTTCCTGCCCCACGTTCATGTACTGCCTTGGGGCCAGCCAGGACTTTGAGGT 2969
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 CATTGCAGCCCGGCTTTTCCTGCCCCACGTTCATGTACTGCCTTGGGGCCAGCCAGGACTTTGAGGT  660

Qy  2970 GGTGCAGAACGCGCCCTTCCGGGAACCTGCCTGAGTACGCTCTCCTTGTGCCCAAAGACACCGCGGC 3029
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 GGTGCAGAACGCGCCCTTCCGGGAACCTGCCTGAGTACGCTCTCCTTGTGCCCAAAGACACCGCGGC  720

Qy  3030 TGTGGCTGCTTCCACCGGGGCC-TTCCTGGTGGGCCANGTACGGCCGCCCAGCCCCAGGCGCGC 3088
         |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   721 TGTGGCTGCTCCACCGGGGGCCTTCCTGGTGGGCCANGTACGGCCGCCCAGCCCCAGGCGCGC  780

Qy  3089 AGATCTCCGAGGCCCTGAGGATCCCAGAATATGGAGGCGCCGTGAGTGTGACCCTGTGGCAGCC 3148
         ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   781 AGATCTCCGAGGCCCTGAGGATCCCAGATATGGAGGCGCCGTGAGTGTGACCCTGTGGCAGCC  840

Qy  3149 GACCCCCCTCCCAAGCCCGCCCGCGTCCCCGGGGGGGAATCCTGAGGCAAAGCCCAGA 3208
         |||||||| ||| ||||||  |||| |||||||||||||  |||||||||||||||||
Db   841 GACCCCCTCCAAGCCCTCATTCACCGGGGGGATCCCGGGGTTGAGGCAAAGCCCAGA  900

Qy  3209 AGCGTGGGGCGTTGCTGGTGCTGTCGTCTCCGGAGGAGGTGAGGGGCGGGGCCCTGAGGCCCAG 3268
         ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AGCGTGGGGCGTTGCTGGTGCTGTCGTCTCCGGAGGAGGTGAGGGGCGGGGCCCTGAGGCCCAG  960

Qy  3269 GCCCAGGAGCGAATACTCCGAGCCCTGGGGGTGGCTTCCGGGCCCTGGCCATCAGGGGGCAGGGGGC 3328
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GCCCAGGAGCGAATACTCCGAGCCCTGGGGGTGGCTGCTGGCCCTGGCCATCAGGGGGCAGGGGGC 1020

Qy  3329 CGTCCAGCAAGCCCTCATTCAACCTTCTGGGGGCCACAGCCCCTGCCGGAGCGGGCCGGATCCC 3388
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 CGTTCCAGCAAGCCCTCATTCACCTTCTGGGGCCACAGCCCCTGCCGGAGCGGGCGGGATCCC 1080

Qy  3389 CCCGGCCATGGCCCTGGGCGGGTTTTGAATGAAAACGACCTGAACTGTCAA 3437
         |||          ||||||||||||||||||||||||||||||||||||||
Db  1081 CGCGGTACATGGCTGGGCTGGGCTGGTTTTGAATG-AACGACCTGAACTGTCAA 1128
```

RESULT 9
AC102953/c

```
LOCUS       AC102953              148996 bp    DNA     linear   PRI 01-JUN-2002
DEFINITION  Homo sapiens BAC clone RP11-1246C19 from 7, complete sequence.
ACCESSION   AC102953
VERSION     AC102953.5  GI:21307577
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148996)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  PUBMED    9847074
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 148996)
  AUTHORS   Du,H. and Kozlowicz,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-1246C19
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 148996)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 148996)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 148996)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 148996)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 148996)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Jun 1, 2002 this sequence version replaced gi:19774660.
COMMENT     Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            --------------------------------------------------------------
            Summary Statistics
            Center project name: H_NH1246C19
            --------------------------------------------------------------

NOTICE:  This sequence may not represent the entire insert of this
         clone. It may be shorter because we only sequence overlapping
         clone sections once, or longer because we provide a small overlap
         between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

-----------------------------

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director) and the Washington University Genome
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GTB/CHR7/, or see http://genome.wustl.edu/gsc
mailto:egreen@nhgri.nih.gov, or send

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,C.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-369D24, 2000 bp overlap;
the clone sequenced to the right is RP11-16P10, 2000 bp overlap.
Actual start of this clone is at base position 174583 of
RP11-369D24; actual end is at base position 63205 of RP11-16P10.
```

FEATURES   Location/Qualifiers
source   1..148996
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="7"
                /map="7"
                /clone="RP11-1246C19"
                /clone_lib="RPCI-11"

Sequence derived from one plasmid subclone, base position 32899 to 33087.

Unresolved tandem repeat from base position 99549 to 100076.

repeat_region   585..893   /rpt_family="Alu"
repeat_region   2518..2618   /rpt_family="MIR"
repeat_region   2708..2769   /rpt_family="MIR"
repeat_region   2873..3034   /rpt_family="MIR"
repeat_region   3347..3438   /rpt_family="MIR"
repeat_region   3439..3574   /rpt_family="L1"
repeat_region   3575..3857   /rpt_family="Alu"
repeat_region   3858..4034   /rpt_family="Alu"
repeat_region   4035..4628   /rpt_family="Alu"
repeat_region   4629..4714   /rpt_family="L1"
repeat_region   4796..5037   /rpt_family="Alu"
repeat_region   5039..5296   /rpt_family="Alu"
repeat_region   6603..6933   /rpt_family="L1"
repeat_region   6938..7061   /rpt_family="Alu"
repeat_region   7145..7243   /rpt_family="L1"
repeat_region   7256..7559   /rpt_family="MER1_type"
repeat_region   9418..9677   /rpt_family="Alu"
repeat_region   9758..9876   /rpt_family="L2"
repeat_region   9886..9926   /rpt_family="MER1_type"
repeat_region   9927..10207   /rpt_family="(TTCC)n"
repeat_region   10208..10220   /rpt_family="Alu"
repeat_region   10227..10366   /rpt_family="MER1_type"
repeat_region   11484..11545   /rpt_family="L2"
repeat_region   12151..12259   /rpt_family="MIR"
repeat_region   12792..13111   /rpt_family="MIR"
repeat_region   14332..14451   /rpt_family="MIR"
repeat_region   14794..14822   /rpt_family="MIR"
repeat_region   14828..15133   /rpt_family="AT_rich"
repeat_region   15194..15288   /rpt_family="Alu"
repeat_region   16148..16439   /rpt_family="MIR"
repeat_region   16468..16637   /rpt_family="Alu"
repeat_region   16647..16944   /rpt_family="Alu"
repeat_region   16972..17244   /rpt_family="Alu"
repeat_region   17254..17442   /rpt_family="L1"
repeat_region   19130..19311   /rpt_family="L2"
repeat_region   19432..19683   /rpt_family="MaLR"
repeat_region   19833..19938   /rpt_family="L2"
repeat_region   20105..20243   /rpt_family="MIR"
repeat_region   20951..21003   /rpt_family="L1"
repeat_region   21298..21415   /rpt_family="L1"
repeat_region   21589..21644   /rpt_family="MIR"
repeat_region   21944..22026   /rpt_family="MIR"
repeat_region   22180..22292   /rpt_family="MIR"
misc_feature   22268..23087
                /note="CpG_island (%GC=72.4, O/e=0.80, #CpGs=75)"
repeat_region   22660..22845   /rpt_family="(CACCC)n"
repeat_region   23225..23354   /rpt_family="MIR"

Query Match   12.2%; Score 418; DB 9; Length 148996;
Best Local Similarity 100.0%; Pred. No. 1.4e-50;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   3018 AGAGCAGGCAGCGGCTGTGCGTTCCTCCACCGGGCGCCTTCCTTGGTGTTGGGCANTGATAGTACGGCCCAGATGGA 3077
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 124110 AGAGCAGGCAGCGGCTGTGCGTTCCTCCACCGGGCGCCTTCCTTGGTGTTGGGCANTGATAGTACGGCCCAGATGGA 124051

Qy   3078 CCCCAGCGCGCAGANTCTCCGAGGCCGCCCTGAGGATCCTGCCATATGAGAGGCCGTGATGTGAGC 3137
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 124050 CCCCAGCGCGCGAGATCTCCGAGGCCGCCCTGAGGATCCTGCCATATGAGAGGCCGTGATGTGAGC 123991

Qy   3138 CTGTGGCAGCCGACCCCCTCCCAAGCCCCGGCCCGTCCCGTCCCCGGGGATCCTCGAGGC 3197
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 123990 CTGTGGCAGCCGACCCCCTCCCAAGCCCCGGCCCGTCCCGTCCCCGGGGATCCTCGAGGC 123931

Qy   3198 AAAGCCCAGGAGAAGGGTGGGCGTTGCTGGCGTTGCTGGCGTCCGAGGAGGTCAGGGCGCCGGCGCCCTG 3257
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 123930 AAAGCCCAGGAGAAGGGTGGGCGTTGCTGGCGTTGCTGGCGTCCGAGGAGGTCAGGGCGCCGGCGCCCTG 123871

Qy   3258 AGGCCAGGCAGGCCCGGAGCAAATACTCCGAGCCCTGGGGGTTGGGGGTTGGGCTCCGGGCCCGGCCCTG 3317
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 123870 AGGCCAGGCAGGCCCGGAGCAAATACTCCGAGCCCTGGGGGTTGGGGGTTGGGCTCCGGGCCCGGCCCTG 123811

Qy   3318 GCAATCAGGGGCCGTTCCAGCAAGCCCTCAATTCACCTTCTGGGCCACAGCGCCTTGCCGCCGGAG 3377
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 123810 GCAATCAGGGGCCGTTCCAGCAAGCCCTCAATTCACCTTCTGGGCCACAGCGCCTTGCCGCCGGAG 123751

Qy   3378 CGGCGGGATCCCCCCGGGCCATGGGCCTGGGCCTGGGTTGGTTTTGGAATTGAAAACGACCTGAAACTGTC 3435
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 123750 CGGCGGGATCCCCCCGGGCCATGGGCCTGGGCCTGGGTTGGTTTTGGAATTGAAAACGACCTGAAACTGTC 123693

RESULT 10
AC120290/c
LOCUS   AC120290   144097 bp   DNA   linear   HTG   23-JUL-2002
DEFINITION Rattus norvegicus clone CH230-101I10, ***, 72 unordered pieces.
ACCESSION   AC120290
VERSION   AC120290.2   GI:21902868

```
KEYWORDS     HTG; HTGS_PHASE1.
SOURCE       Rattus norvegicus.
ORGANISM     Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE    1  (bases 1 to 144097)
AUTHORS      Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
             Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
             Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
             Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
             Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
             Carron,T.F., Carter,M. Cavazos,S.R., Chacko,J., Chavez,D.,
             Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
             Cleveland,C.D., Cox,C., Coyle,M.D. Dathorne,S.R., David,R.,
             Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
             Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
             Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
             Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
             Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
             Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
             Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
             Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
             Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
             Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
             Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
             Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
             Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
             Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
             Lozado,R.J., Lu,J., Lucier,A., Lucier,R., Luna,R., Ma,J.
             Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
             Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
             Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
             Moser,M., Neal,D., Newton,J., Nguyen,N., Nguyen,A., Nguyen,N.,
             Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
             Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
             Peters,L., Pickens,R., Primus,E., Pu,L.L., Puazo,M., Quiles,M.,
             Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
             Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
             Soderlund,S., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
             Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
             Tansey,J., Taylor,C., Taylor,J., Telford,B., Thomas,N., Thomas,S.,
             Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
             Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
             Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
             Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
             Weinstock,G. and Gibbs,R.
TITLE        Direct Submission
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 144097)
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (05-MAY-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE    3  (bases 1 to 144097)
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Jul 18, 2002 this sequence version replaced gi:20452937.
             --------- Genome Center
             Center: Baylor College of Medicine
             Center code: BCM
             Web site: http://www.hgsc.bcm.tmc.edu/
             Contact: hgsc-help@bcm.tmc.edu
             --------- Project Information
             Center project name: GWBB
             Center clone name: CH230-101110
             --------- Summary Statistics
             Sequencing vector: plasmid;
             Chemistry: Dye-terminator Big Dye; 100% of reads
```

```
* NOTE: Estimated insert size may differ from sequence length
*       (see http://www.hgsc.bcm.tmc.edu/docs/genbank.draft.data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 72 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

Assembly program: Phrap; version 0.990329
Consensus quality: 62007 bases at least Q40
Consensus quality: 68326 bases at least Q30
Consensus quality: 73396 bases at least Q20
```

```
      1:    1060: contig of 1060 bp in length
   1061:    1160: gap of unknown length
   1161:    2270: contig of 1110 bp in length
   2271:    2370: gap of unknown length
   2371:    3371: contig of 1001 bp in length
   3372:    3471: gap of unknown length
   3472:    4834: contig of 1363 bp in length
   4835:    4934: gap of unknown length
   4935:    5992: contig of 1058 bp in length
   5993:    6092: gap of unknown length
   6093:    7670: contig of 1578 bp in length
   7671:    7770: gap of unknown length
   7771:    9212: contig of 1442 bp in length
   9213:    9312: gap of unknown length
   9313:   10541: contig of 1229 bp in length
  10542:   10641: gap of unknown length
  10642:   12072: contig of 1431 bp in length
  12073:   12172: gap of unknown length
  12173:   13552: contig of 1380 bp in length
  13553:   13652: gap of unknown length
  13653:   15095: contig of 1444 bp in length
  15096:   15196: gap of unknown length
  15197:   16222: contig of 1026 bp in length
  16223:   16322: gap of unknown length
  16323:   17655: contig of 1333 bp in length
  17656:   17755: gap of unknown length
  17756:   18939: contig of 1184 bp in length
  18940:   19039: gap of unknown length
  19040:   20566: contig of 1527 bp in length
  20567:   20666: gap of unknown length
  20667:   21987: contig of 1321 bp in length
  21988:   22087: gap of unknown length
  22088:   23119: contig of 1032 bp in length
  23120:   23219: gap of unknown length
  23220:   24254: contig of 1035 bp in length
  24255:   24354: gap of unknown length
  24355:   25397: contig of 1043 bp in length
  25398:   25497: gap of unknown length
  25498:   27045: contig of 1548 bp in length
  27046:   27145: gap of unknown length
  27146:   28578: contig of 1433 bp in length
  28579:   28678: gap of unknown length
  28679:   29972: contig of 1294 bp in length
  29973:   30072: gap of unknown length
  30073:   31567: contig of 1495 bp in length
  31568:   31667: gap of unknown length
  31668:   33101: contig of 1434 bp in length
  33102:   33201: gap of unknown length
  33202:   34444: contig of 1243 bp in length
  34445:   34544: gap of unknown length
  34545:   35824: contig of 1280 bp in length
  35825:   35924: gap of unknown length
  35925:   37318: contig of 1394 bp in length
  37319:   37418: gap of unknown length
  37419:   38684: contig of 1266 bp in length
  38685:   38784: gap of unknown length
  38785:   40419: contig of 1635 bp in length
  40420:   40519: gap of unknown length
```

```
*    40520    42463: contig of 1944 bp in length
*    42464    42553: gap of unknown length
*    42564    44280: contig of 1717 bp in length
*    44281    44280: gap of unknown length
*    44281    45965: contig of 1585 bp in length
*    45966    46065: gap of unknown length
*    46066    47457: contig of 1392 bp in length
*    47458    47557: gap of unknown length
*    47558    49418: contig of 1861 bp in length
*    49419    49518: gap of unknown length
*    49519    50846: contig of 1328 bp in length
*    50847    50946: gap of unknown length
*    50947    53079: contig of 2133 bp in length
*    53080    53179: gap of unknown length
*    53180    54796: contig of 1617 bp in length
*    54797    54896: gap of unknown length
*    54897    55980: contig of 1084 bp in length
*    55981    56080: gap of unknown length
*    56081    58602: contig of 2522 bp in length
*    58603    58702: gap of unknown length
*    58703    60423: contig of 1721 bp in length
*    60424    60523: gap of unknown length
*    60524    62328: contig of 1805 bp in length
*    62329    62428: gap of unknown length
*    62429    64486: contig of 2058 bp in length
*    64487    64586: gap of unknown length
*    64587    66659: contig of 2073 bp in length
*    66660    66759: gap of unknown length
*    66760    68152: contig of 1393 bp in length
*    68153    68252: gap of unknown length
*    68253    70718: contig of 2466 bp in length
*    70719    70818: gap of unknown length
*    70819    72273: contig of 1455 bp in length
*    72274    72373: gap of unknown length
*    72374    74454: contig of 2081 bp in length
*    74455    74554: gap of unknown length
*    74555    76795: contig of 2241 bp in length
*    76796    76895: gap of unknown length
*    76896    79006: contig of 2111 bp in length
*    79007    79106: gap of unknown length
*    79107    82084: contig of 2978 bp in length
*    82085    82184: gap of unknown length
*    82185    83489: contig of 1305 bp in length
*    83490    83589: gap of unknown length
*    83590    86730: contig of 3141 bp in length
*    86731    86830: gap of unknown length
*    86831    88726: contig of 1896 bp in length
*    88727    88826: gap of unknown length
*    88827    90953: contig of 2127 bp in length
```

```
Query Match     5.7%;  Score 194.4;  DB 2;  Length 144097;
Best Local Similarity   84.2%;
Matches  219;  Conservative   0;  Mismatches  41;  Indels   0;  Gaps   0;
                            Pred. No. 9.3e-19;

Qy    632 TCCAGATTTTCCCGCTCAGCCCGGACCCTCGGTGGCAGAGCTTCCAGTCCCCGCCCCGTGG 691
          |||||| ||| ||||||||||  ||||||||||||| |||||||||||||||||||||||
Db  51563 TCCAGATCTTCCCACTGAGCCCGGATCCACGGTGGCAGAGCTCCAGTCCCGCGCCCCGTGG 51504

Qy    692 CCCTCGCCCTGCAGCAGGCCCCTGGGGCCCAGGAGCTTGGCCCCGGCTGCGTCCAGGGCAG 751
          |||| ||   ||||||||| ||||||||| |||||||||||||||  |||||||||||
Db  51503 CCCTGGCACTGCAGCAAGCCCTGGGCCCAAGAGTTGGCCGGTGCGTCCGGGCCAGGGCAGG 51444

Qy    752 AGGTGCCGGGGATCACGGTGCGTGTCCTGCAGGGCCCTGCACCCTGCTCAGCTCCCCAC 811
          |||||||||| |||||  |||||| |||||||| |||||||||||||| |||||||
Db  51443 AGGTGCCCGGGCATCACAGTTCGCCTTCTGCAGGGCCCATGCACCCTGCTGAGCTCCCCGC 51384

Qy    812 ACGGCGGGTGCCCTGGTGATGTCCATGCACCGTAGCCACTTCCTGCCGCCCGGCCGGCTGC 871
          |||| || || |||||||||||||||||||||||||||| |||||| ||||||||| ||
Db  51383 ACGGGGGTGACCCTTGCTGGCTATGCACCACACACCACTTCCTGTGCTCCGCCCTTCGATGC 51324

Qy    872 GCCAGCTGCGCCAGTACCAG 891
          ||||||||||| ||||||||
Db  51323 GCCAGCTGCGCCAGTACCAG 51304
```

```
COMMENT     -------------- Genome Center --------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information --------------
            Center project name: M_BA0168E11
            -------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *      1    2186: contig of 2186 bp in length
            *   2187    2286: gap of unknown length
            *   2287    6318: contig of 4032 bp in length
            *   6319    6418: gap of unknown length
            *   6419   33825: contig of 27407 bp in length
            *  33826   33925: gap of unknown length
            *  33926   80048: contig of 46123 bp in length
            *  80049   80148: gap of unknown length
            *  80149  128323: contig of 48175 bp in length
            * 128324  128423: gap of unknown length
            * 128424  219964: contig of 91541 bp in length
            -------------- Summary Statistics --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 217436 bases at least Q40
            Consensus quality: 217966 bases at least Q30
            Consensus quality: 218400 bases at least Q20
FEATURES             Location/Qualifiers
     source          1..219964
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-168E11"
     misc_feature    1..2186
                     /note="assembly_name:Contig13"
     misc_feature    2287..6318
                     /note="assembly_name:Contig14"
     misc_feature    6419..33825
                     /note="assembly_name:Contig15"
```

RESULT 12

```
misc_feature    33926..80048
                /note="assembly_name:Contig16"
misc_feature    80149..128323
                /note="assembly_name:Contig17"
misc_feature    128424..219964
                /note="assembly_name:Contig18"
BASE COUNT    51706 a  56458 c  58030 g  53268 t  502 others
ORIGIN

Query Match    5.4%;  Score 186.2;  DB 2;  Length 219964;
Best Local Similarity   80.48%;  Pred. No. 1.3e-17;
Matches 218;  Conservative 0;  Mismatches 53;  Indels 0;  Gaps 0;

Qy    631 CTCCAGATTTTCCGCCTCAGCCCGGACCCCTCCAGTCCCCGCCCGTG 690
          ||||||| |||| | || || ||||  ||  |||||||| |||| |||
Db 157304 CTCCAGATCTTCCCACTGGAGCCGCGGCGGAGGCTCCAGCCCCCCTT 157363

Qy    691 GCCCTGCCCCTGCAGCAGGCAGGAAGCTGGCCCGCCGCGTCGTCAGGGCAGCAGCCCC 750
          ||||| || ||||||||| |||  | |||||||    | || | | ||  |||  |||
Db 157364 GCCCTGGCACTGCAGCAGCAAGCCCCTGGGGCCCAAGAGTGGCTGCTGTGTCCCGCCAGGGAACCC 157423

Qy    751 GAGGTGCCGGGCAATCACGGGTGCGTGTCCTGCAGGCCCCCGCCCACCCTGCCCGCCCA 810
          |||||||| ||||||||| |||| | ||| || ||| |||||||||| |||||||||||
Db 157424 GAGGTGCCAGGCAATCACAGTGCCGTCCGCCTTCTGCAGGCCCCCACCCACCCTGCCCGCCCA 157483

Qy    811 CACGCGCGGGGTGCCCTGGTTGAGTGCCATGCACCGCACTTCCTGGCCTGCCCTGCTG 870
          |||| ||  ||| | |||  | |  |||  | |||  ||||||||||| || ||||| ||
Db 157484 CACGGTGGGGCACCCTTGCTCTGGCTATGCAACACAGGCACTTCCTTCCTGCCCCACTGATG 157543

Qy    871 CGGCCAGCGTCTACCAGTACCAGCGCGGTGTGTGC 901
          |||||||| |||| || ||||||||| |||||||
Db 157544 CGCCAGCGTCTACCAGTACCAGGTATGCTGGC 157574
```

ACCESSION / LOCUS block

```
LOCUS       AC004642              148432 bp    DNA     linear   INV 01-MAY-1998
DEFINITION  Drosophila melanogaster DNA sequence (P1s DS00543 (D193) and
            DS02867 (D200)), complete sequence.
ACCESSION   AC004642 AC004302 AC003551 AC003552 AC003553 AC003554 AC003555
            AC003556 AC003557 AC003558 AC004308 AC003848 AC003849
            AC003576 AC003577 AC003578 AC003579 AC003580 AC003581 AC003582
            AC003583 AC003584 AC003585 AC003850 AC003586 AC003587 AC003588
            AC003589 AC003590 AC003591 AC003592
VERSION     AC004642.1  GI:3097827
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster
ORGANISM    Drosophila melanogaster (subclones in Sac from P1 clones DS00543
            (D193) and DS02867 (D200)) DNA.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 148432)
AUTHORS     Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
            Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Shir,E.,
            Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L.
            and Kimmel,B.
TITLE       Sequencing of Drosophila chromosome 2R, region 60B1-60B10
JOURNAL     Unpublished (1997)
REFERENCE   2  (bases 1 to 148432)
AUTHORS     Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
            Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Shir,E.,
            Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L.
            and Kimmel,B.
TITLE       Direct Submission
JOURNAL     Submitted (01-MAY-1998) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
```

RESULT 13

```
COMMENT     Berkeley, CA 94720, US
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site
            (http://fruitfly.berkeley.edu/sequence-archive.html) or send email
            to drosophila@mbgc.lbl.gov.
            Library locations: 63-6, 83-30.
FEATURES             Location/Qualifiers
     source          1..148432
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="60B1-60B10"
                     /chromosome="2R"
                     /clone="P1s DS00543 (D193) and DS02867 (D200)"
                     /note="d193 extends from bp 1 to bp 83271 and d200 extends
                     from bp 81802 to bp 148432."
BASE COUNT    38713 a  34662 c  35252 g  39805 t
ORIGIN

Query Match    3.1%;  Score 107.2;  DB 3;  Length 148432;
Best Local Similarity   60.18%;  Pred. No. 2.4e-06;
Matches 197;  Conservative 0;  Mismatches 128;  Indels 3;  Gaps 1;

Qy    131 TGGACATCTGGTTTCCGGAGGAGAAGCCACTCGCCACCGCCTTCCTGTGTGGACA---CAT 187
          |||||| | ||| || |||||  |||| | ||| || ||||  ||| || || |   |||
Db   5294 TGGACTATTGGTTTCCGCCCGGAGGAGACCAGCAGCGCCTTCGCCTTTCTGCCCAGCATGCGGC 187

Qy    188 CGGAGGAGGCGCGTGCTGCTTCCCTGACTGGCCTGGGAAGCTGCGGCAATGGACCATCCGTTTCCTGAGGTGC 247
          ||||| ||| | | || |  || ||||  ||| |||||||| || | ||| ||    ||   |||| |||||||
Db  50234 AGGACGCAGGTGCAACTCGCTGCCCACACTGGCCTGAAGCTGAAGACTGAAGATCGGCGCTCACTGGGTG 50175

Qy    248 TCCGCCTGGTGGACGCGCCCGTTCGAGGACCTGGAGCCTGAGCTGCTGCTTGTTGTGGC 307
          ||| || |||| ||||  ||| |  |||| | |||   |  | |||  |  || | | ||
Db  50174 ACAGACTAATTGAGGCAGCTGTTCAACAGATCTAACGCCGGATCAGAATCGTTCGTTTGTGC 50115

Qy    308 AGTCGTTTGGGCATCCCCGTGTTCCAGCACAGAGCAAACTCCTCCACGTTCCTGGACCAGGCGAG 367
          | ||| ||||| | | || || |||||  ||| ||  | ||  | | |||| |||| |||| |||
Db  50114 AAAACTTTGGCACGCGCGGGTTCGAACTCCGATGTCCAACAGGTCCGCGGCAATGCTGGACACCGCTG 50055

Qy    368 TGGCCCAGGACCCCCAGATCTCTGGAGCAGGAGACATCATTGGACAAGAGATTACATGGCCCACC 427
          ||| || | | ||| |||||| | |||||  ||||||  |  ||| |  |    ||| | ||||||
Db  50054 TACTGGAGCAGTTTGGATCTGGTGAAGAATGCCCATTCTGAACAAGGGCCTAATCTGGCCCAGC 49995

Qy    428 TGGGTGGAGGTCCTCAGCATGAGCCGCGGCGC 455
          ||| | ||| |  | || || |||  |||| |
Db  49994 TAATCGAGATTCAGCAGGCGCGGGGGGGGTGC 49967
```

```
LOCUS       AC099018              154840 bp    DNA     linear   INV 08-NOV-2001
DEFINITION  Drosophila melanogaster, chromosome 2R, region 60A-60B, BAC clone
            BACR28M05, complete sequence.
ACCESSION   AC099018
VERSION     AC099018.1  GI:16798950
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 154840)
AUTHORS     Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Wan,K.H.,
            Rogers,Y., An,H., Baldwin,D., Ballew,R.M., Bang01,R.C.,
            Carlson,J.W., Center,A., Champe,M., Banzon,R.Y., Busam,D.A.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
```

RESULT 14

```
AUTHORS    Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
           McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nunoo,J.,
           Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
           Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S.,
           Stapleton,M., Strong,R., Sviraskas,R., Tector,C., Williams,F.,
           Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE    Sequencing of Drosophila chromosome 2R, region 60A-60B
  JOURNAL  Unpublished
REFERENCE  2  (bases 1 to 154840)
  AUTHORS  Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
           Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
           Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
           Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
           Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
           Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
           Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
           Ibegwam,C., Jalali,M., Li,P., Mattei,B., Moshrefi,A.,
           McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
           Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
           Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
           Stapleton,M., Strong,R., Sviraskas,R., Tector,C., Williams,S.M.,
           Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE    Direct Submission
  JOURNAL  Submitted (08-NOV-2001) Berkeley Drosophila Genome Project, MS
           64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
           Berkeley, CA 94720, US
COMMENT    Sequence submitted by:
           Berkeley Drosophila Genome Project
           Lawrence Berkeley National Laboratory, MS 64-121
           Berkeley, CA 94720
           This sequence was assembled using end sequences from a whole genome
           shotgun and from subclones of this sequence, including its location
           For further information about this sequence, including its location
           and relationship to other sequences, please visit our sequence
           archive Web site (http://www.fruitfly.org/sequence/) or send email
           to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1..154840
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="60A-60B"
                     /clone="BACR28M05 (D1305)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe.6)"
BASE COUNT    41075 a   36234 c   35932 g   41599 t
ORIGIN
```

```
Query Match      3.1%;  Score 107.2;  DB 3;  Length 154840;
Best Local Similarity   60.1%;  Pred. No. 2.4e-06;
Matches 197;  Conservative   0;  Mismatches 128;  Indels   3;  Gaps   1;

Qy    131 TGGACAATCTGGTTTCCGGAGGAGAAGCCACTGCCCACCGCCTTCCTGGTGGACA---CAT 187
          ||||  ||| ||||||||||||  ||||   |||||||||||||||||  ||||   |||
Db 132928 TGGACTAATTGGTTTCCGGCCCGGAGGAGGACCAGCCGCCGGCCTTTCTGCCCAGCATGCC 132987

Qy    188 CGGAGGAGCCGCGCTGCTTCCTCGCTGGCTGGAGCTCGGCATGATACCGTTCTGAGGTGCC 247
          |||||||||||||||||| ||||||||||| |||| |||||| ||||| ||||||||||||
Db 132988 AGGAGCAGGTGCAACTGCTGCCCCGACTGGGCTGAAGCTGAAGATGACCCGCTCATCGGTGG 133047

Qy    248 TCCGCCTGGTGGACGCCGCCCTCCAGGAACCTGGAGCCGCAGTCGCTGCTGCTTCGGTGC 307
          ||| || |||| ||| ||||| ||||| ||||||||| |||| ||| || |||| ||||
Db 133048 ACAGACTAATTGAGGCAGTTCTCAAGCATCTACCGCGGATCAGAATCGTGCCTTTGTGC 133107

Qy    308 AGTCGTTTGGCATCCCGTGTCCACCGACGATGAGCAGAAACTCCTCCAGTTCCTGGAACAG 367
          || | ||||| |||||||||||||||| || ||||| | ||||||| |||| |||||||||
Db 133108 AAAACTTTGGCACGCGGTCAACTCGATTGTCCAAGCTGCTGGACACCGCTG 133167

Qy    368 TGGCCCACGACCCCCGTTGATCTGGTGAAGAATGCCATTTCTGAACAAGGCCCACC 427
          |||||| ||| |||||||| ||||||||||||||| |||||| || |||||||
Db 133168 TACTGGAGCAGTTTGATCTGGTGAAGAATGCCATTTCTGAACAAGGCCCTATCTCGGCCCAGC 133227
```

RESULT 14 (continued)

```
Qy    428 TGGTGGAGGTCCAGCCATGAGCGCGCGGCGC 455
          | ||| ||| || ||||||||| |||| ||
Db 133228 TAATCGAGAATTCAGCAGGCGCGGCGGGTGC 133255
```

RESULT 15

```
LOCUS       AC020509                157851 bp    DNA     linear   HTG 03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC020509
VERSION     AC020509.1  GI:6664388
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 157851)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10213488 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            1..157851
FEATURES             Location/Qualifiers
     source          1..157851
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    41166 a   36910 c   37402 g   42373 t
ORIGIN
```

```
Query Match      3.1%;  Score 107.2;  DB 2;  Length 157851;
Best Local Similarity   60.1%;  Pred. No. 2.4e-06;
Matches 197;  Conservative   0;  Mismatches 128;  Indels   3;  Gaps   1;

Qy    131 TGGACAATCTGGTTTCCGGAGGAGAAGCCACTGCCCACCGCCTTCCTGGTGGACA---CAT 187
          ||||  ||| ||||||||||||  ||||   |||||||||||||||||  ||||   |||
Db 82298 TGGACTAATTGGTTTCCGGCCCGGAGGAGGACCAGCCGCCGGCCTTTCTGCCCAGCATGCC 82239

Qy    188 CGGAGGAGCCGCGCTGCTTCCTCGCTGGCTGGAGCTCGGCATGATACCGTTCTGAGGTGCC 247
          |||||||||||||||||| ||||||||||| |||| |||||| ||||| ||||||||||||
Db 82238 AGGAGCAGGTGCAACTGCTGCCCCGACTGGGCTGAAGCTGAAGATGACCCGCTCATCGGTGG 82179

Qy    248 TCCGCCTGGTGGACGCCGCCCTCCAGGAACCTGGAGCCGCAGTCGCTGCTGCTTCGGTGC 307
          ||| || |||| ||| ||||| ||||| ||||||||| |||| ||| || |||| ||||
Db 82178 ACAGACTAATTGAGGCAGTTCTCAAGCATCTACCGCGGATCAGAATCGTGCCTTTGTGC 82119

Qy    308 AGTCGTTTGGCATCCCGTGTCCACCGACGATGAGCAGAAACTCCTCCAGTTCCTGGAACAG 367
          || | ||||| |||||||||||||||| || ||||| | ||||||| |||| |||||||||
Db 82118 AAAACTTTGGCACGCGGTCAACTCGATTGTCCAAGCTGCTGGACACCGCTG 82059

Qy    368 TGGCCCACGACCCCCGTTGATCTGGTGAAGAATGCCATTTCTGAACAAGGCCCACC 427
          |||||| ||| |||||||| ||||||||||||||| |||||| || |||||||
Db 82058 TACTGGAGCAGTTTGATCTGGTGAAGAATGCCATTTCTGAACAAGGCCCTATCTCGGCCCAGC 81999

Qy    428 TGGTGGAGGTCCAGCCATGAGCGCGCGGCGC 455
          | ||| ||| || ||||||||| |||| ||
Db 81998 TAATCGAGAATTCAGCAGGCGCGGCGGGGTGC 81971
```

```
LOCUS       AE003462               303823 bp    DNA     linear   INV 28-JUN-2002
DEFINITION  Drosophila melanogaster 2R section 70 of 74 of the complete arm,
            complete sequence.
ACCESSION   AE003462
VERSION     AE003462.2  AE013599 AE002575
            AE003462.2  GI:21626662
```

```
KEYWORDS   .
SOURCE     fruit fly.
  ORGANISM Drosophila melanogaster
           Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
           Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
           Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1  (bases 1 to 303823)
  AUTHORS  Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
           Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
           George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
           Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
           Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
           Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
           Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
           Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
           Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
           Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
           Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
           Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
           Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
           Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
           Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
           Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
           Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
           Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
           Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
           Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
           Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
           Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
           Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
           Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
           McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
           Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
           Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
           Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
           Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
           Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
           Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
           Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
           Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
           Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
           Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
           Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
           Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
           Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE    The genome sequence of Drosophila melanogaster
  JOURNAL  Science 287 (5461), 2185-2195 (2000)
  PUBMED   10731132
  MEDLINE  20196006
REFERENCE  2  (bases 1 to 303823)
  AUTHORS  Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D.,
           Banzon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A.,
           Champe,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V.,
           Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferriera,S., Frise,E.,
           Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houck,J.,
           Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M.,
           Kruse,D., Li,P., Mattei,B., McIntosh,T.C., Moy,M.,
           Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacleb,J., Paragas,V.,
           Park,S., Patel,S., Pfeiffer,B., Phouanenavong,S., Pittman,G.S.,
           Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R.,
           Svirskas,R., Tector,C., Tyler,D., Williams,S.M., Zaveri,J.S.,
           Smith,H.O., Venter,J.C. and Rubin,G.M.
  TITLE    Sequencing of Drosophila melanogaster genome
  JOURNAL  Unpublished
REFERENCE  3  (bases 1 to 303823)
  AUTHORS  Misra,S., Crosby,M.A., Matthews,B.B., Bayraktaroglu,L.,
           Campbell,K., Hradecky,P., Huang,Y., Kaminker,J.S., Prochnik,S.E.,
           Smith,C.D., Tupy,J.L., Bergman,C.M., Berman,B.P., Carlson,J.W.,
           Celniker,S.E., Clamp,M.E., Drysdale,R.A., Emmert,D., Frise,E., de
           Grey,A.D.N.J., Harris,N.L., Harris,N.L., Kronmiller,B., Marshall,B.,
           Milburn,G.H., Richter,J., Russo,S., Searle,S.M.J., Smith,E.,
           Shu,S., Smutniak,F., Whitfield,E.J., Yamada,C., Ashburner,M.,
           Gelbart,W.M., Rubin,G.M., Mungall,C.J. and Lewis,S.E.
REFERENCE  4  (bases 1 to 303823)
  AUTHORS  Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE    Annotation of Drosophila melanogaster genome
  JOURNAL  Unpublished
REFERENCE  5  (bases 1 to 303823)
  AUTHORS  Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE    Direct Submission
  JOURNAL  Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
           Rockville, MD, USA
REFERENCE  5  (bases 1 to 303823)
  AUTHORS  Lewis,S.E.
  TITLE    Direct Submission
  JOURNAL  Submitted (31-MAY-2002) University of California Berkeley, 539 Life
           Sciences Addition, Berkeley, CA 94720, USA
COMMENT    On Jun 28, 2002 this sequence version replaced gi:7291637.
FEATURES             Location/Qualifiers
     source          1..303823
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2"
                     /map="2R"
     gene            complement(11873..23611)
                     /gene="CG11169"
                     /map="60A9-60A10"
                     /db_xref="FLYBASE:FBgn0011169"
     mRNA            complement(join(11873..12379,18020..18143,23421..23611))
                     /gene="CG11169"
                     /db_xref="FLYBASE:FBgn0034905"
                     /product="CG11169-RA"
                     /db_xref="FLYBASE:FBan0011169"
     CDS             complement(join(11873..12379,18020..18091))
                     /gene="CG11169"
                     /note="CG11169 gene product"
                     /codon_start=1
                     /product="CG11169-PA"
                     /protein_id="AAF47060.1"
                     /db_xref="GI:7291638"
                     /db_xref="FLYBASE:FBgn0034905"
                     /db_xref="FLYBASE:FBan0011169"
                     /translation="MSSPIVLFTALLLVSSTICQGLPDLLEKQMTPEQLQSVFHVDTHD
                     AVPHYELVQLQHHENNNRRRSIGKGPKVNAALPPHHVKKDLSKNAYYSELKHEAIA
                     SGGNHLFAPEVSAIKSHNVSFSARGSQOHLNLSLRATVQGLFKGAPHQLRMWTVGSEPNAT
                     HGIDLQEIVHEQVSRVYSMSIQLKIIVHRVY"
     gene            complement(16742..17216)
                     /gene="CG15800"
                     /map="60A10-60A10"
                     /db_xref="FLYBASE:FBgn0034904"
     mRNA            complement(16742..17216)
                     /gene="CG15800"
                     /product="CG15800-RA"
                     /db_xref="FLYBASE:FBan0015800"
                     /db_xref="FLYBASE:FBgn0034904"
     CDS             complement(16743..17216)
                     /gene="CG15800"
                     /note="CG15800 gene product"
                     /codon_start=1
                     /product="CG15800-PA"
                     /protein_id="AAF47061.1"
                     /db_xref="GI:7291639"
                     /db_xref="FLYBASE:FBan0015800"
                     /db_xref="FLYBASE:FBgn0034904"
                     /translation="MVFEETRDGRTISVNLALLEKSLVIREMCRVGQVPEDEDFVLPL
                     HGIHSNVLLKILLWAEYEHEKHEPPAWWDSKDPLPAEEIELQISDMDKEFLRVEIDSIC
                     KIMEGCNYLIDIPWLYKLCAQKLVFLSSREPETQFSGYVEKIPRFQCDHFIFQIOKE"
     gene            60A10-60A10"
                     /gene="CG30177"
                     /map="60A10-60A10"
                     /db_xref="FLYBASE:FBan0030177"
                     /db_xref="FLYBASE:FBgn0050177"
     mRNA            27566..28291
                     /gene="CG30177"
                     /product="CG30177-RA"
                     /db_xref="FLYBASE:FBan0030177"
```

```
CDS             /db_xref="FLYBASE:FBgn0050177"
                27671..28291
                /gene="CG30177"
                /note="CG30177 gene product"
                /codon_start=1
                /product="CG30177-PA"
                /protein_id="AAM68270.1"
                /db_xref="GI:21626663"
                /db_xref="FLYBASE:FBan0030177"
                /db_xref="FLYBASE:FBgn0050177"
                /translation="MLRIVYIYGTLLPGWSSLRAIGHDQRHFIELWLKYWVIIYALLQ
                GLGVLITDFLLGGLSFYAGLKLILSVGLWFSAPYSTNHLFNILNTYSMGKLICPVVEKGV
                RWHEKVCRNIFRGFWKKSPIVTVWMPRDEDEHITEPRINNRLLKRELSNLMTOIAKNSA
                DHRIQHESPMHISPRHLRAAEILDKFNRESMHNVIDIYLGQRGGDQQ"

gene            complement(30339..30720)
                /gene="CG13561"
                /map="60A11-60A11"
                /db_xref="FLYBASE:FBan0013561"
                /db_xref="FLYBASE:FBgn0034906"

mRNA            complement(join(30339..30494,30556..30720))
                /gene="CG13561"
                /product="CG13561-RA"
                /db_xref="FLYBASE:FBan0013561"
                /db_xref="FLYBASE:FBgn0034906"

CDS             complement(join(30339..30494,30556..30720))
                /gene="CG13561"
                /note="CG13561 gene product"
                /codon_start=1
                /product="CG13561-PA"
                /protein_id="AAF47062.1"
                /db_xref="GI:7291640"
                /db_xref="FLYBASE:FBan0013561"
                /db_xref="FLYBASE:FBgn0034906"
                /translation="MRMQLLKKASGYKPFLYNICQSDVCEYLEKRNHPFINIILSSFG
                NRFNVNKCPIPPEIVLEHFRFPVKVLDMPLPFGDYGLFTTFTEHRSELAQVKVFTL
                TEYR"

gene            30835..31912
                /gene="CG5539"
                /map="60A11-60A11"
                /db_xref="FLYBASE:FBan0005539"
                /db_xref="FLYBASE:FBgn0034907"

mRNA            30835..31912
                /gene="CG5539"
                /product="CG5539-RA"
                /db_xref="FLYBASE:FBan0005539"
                /db_xref="FLYBASE:FBgn0034907"

CDS             31014..31853
                /gene="CG5539"
                /note="CG5539 gene product"
```

```
Query Match      3.1%;   Score 107.2; DB 3;    Length 303823;
Best Local Similarity   60.1%;   Pred. No. 2.1e-06;
Matches 197; Conservative    0; Mismatches 128;  Indels    3;  Gaps    1;

Qy    131 TGGACATCTHGGTTTCCGGAGGAGAAGCCACTGCCCACCGCCTTCCTGGGTGGACA---CAT  187
          ||||| |||||||||||                 ||| |||||||||| ||||||||   |||
Db 267562 TGGACTATTGGTTTCCGCCCGGACGACAGGCGCCCGTGCCCTTTCTGCCCAGCATGCGC  267621

Qy    188 CGGAGGAGGCGCTGCTGCTTCCTGACTGGCTGAAGCTGCGGCATGAATCCGTTCTGAGGTGC  247
          ||||| |||| ||| |  ||||||||||||||| ||||| |||||||||||||||||||||
Db 267622 AGGAGCAGGTGCAACTGCTGCCCGACTGGCTGAAGCTGAAGATGAGATCCCGTCATCGCGTGG  267681

Qy    248 TCCCGCCTCGTTGGACGCGCGCCCTCCAGGACCTGGAGCCCGCAGCAGTCTGCCTGGTTCCTGC  307
          | ||| |||||||||||||||||||| |||||||||||| ||||||||||||| || ||||
Db 267682 ACAGACTAAATTGAGGCAGCTCTCAAACGATCTAACGCCGGATCAGCAGTCGTGCTCTTTGTGC  267741

Qy    308 AGTCGTTHGGCATCCCCGTGTCCAGCATGACCAAACTCCTCCAGTTCCTGGACCAGGCAG  367
          ||||||| ||||||| ||||| |||| |||| ||||  ||| | |||| |||||||||||
Db 267742 AAAACTTTGGCACGCGGTCAACTCGAHGTCCAAGCTGCTGCAAATGCTGGACCACCGCTG  267801

Qy    368 TGGCCCACGACCCCAGACTCTGGGACAGCAGAACATCACATGGACAGAATTACATGGCCCACC  427
          ||  ||||| ||||||||||||||| ||  || || |||| ||| ||||||||||||||||
Db 267802 TACTGGAGCAGTTGATCTGGTGCAGAATGCCATTCTGAACAAGGCCTATTCTGGCCCAGC  267861
```

```
Search completed: March 24, 2003, 16:24:37
Job time : 10061 secs

Qy    428 TGGTGGGAGGTCCAGCATGAGCAGCGCGGCGC  455
          | ||| |||| |||||||| | |||| | ||| || ||
Db 267862 TAATCGAGATTCAGCAGCAGCAGGCGCGGGGGTGC  267889
```